1949

Technique for testing the homogeneity of separately-evaluated behavior characteristics

Charles Owen Neidt
Iowa State College

NOTE TO USERS

TECHNIQUE FOR TESTING THE HOMOGENEITY OF SEPARATELY-EVALUATED BEHAVIOR CHARACTERISTICS

by

Charles Owen Neidt

A Dissertation Submitted to the
Graduate Faculty in Partial Fulfillment of
The Requirements for the Degree of

DOCTOR OF PHILOSOPHY

Major Subject: Vocational Education

Approved:

Signature was redacted for privacy.

In Charge of Major Work

Signature was redacted for privacy.

Head of Major Department

Signature was redacted for privacy.

Dean of Graduate College

Iowa State College

1949

TABLE OF CONTENTS

LIST OF TABLES

I. INTRODUCTION

The usual procedure in scoring psychological tests, whether they are measures of aptitudes, attitudes, interests, or personality characteristics, is that of assigning numerical values to a subject's responses to a series of items. Although magnitudes of the values assigned to responses vary greatly from test to test, the value assigned to the response has definite implications for interpretation. For example, the numerical value may mean to perform correctly, to agree, or to exhibit some personality characteristic. The values then are combined in any of several ways such as summation or median determination to obtain a subject's test score. Despite differences in magnitude of the values and differences in methods of combining the values to obtain the test score, the concept of assigning values to the item responses is universal in psychometric practice.

If prediction of performance from scores of psychological tests is to be accurate, consideration must be given to the behavior represented by a subject's pattern of responses. If different types of items in the testing instrument elicit different types of behavior from the subject, then accuracy of measurement is reduced whenever the values assigned to the various responses are combined. For example, if a subject

possesses differing amounts of several behavior characteristics all represented by a single test score, the relative amounts of the characteristics may be obscured in this single score.

Cureton has emphasized this point in the following statement:

> The most important requirement for a test whose scores are to be interpreted as measurements would seem to be that its items all draw upon the same set of abilities and traits.[1]

Dunlap has also stated:

> It seems to the writer a more logical procedure to build up empirically tests which measure a single factor and then determine their validity, than to try to isolate the various factors of our heterogeneous tests.[2]

The need for testing homogeneity is not confined to evaluating psychological characteristics. In educational evaluation, appraisal of instruction is now generally accepted to be satisfactory only when the degree to which pupils have achieved each of the important objectives has been evaluated. The traditional practice of measuring pupil ability to recall information and inferring from the results the totality of pupil achievements has been abandoned in theory if not in practice.

---

[1] E. E. Cureton. Quantitative psychology as a rational science. Psychometrika. 11:191-196. 1946. p. 195.

[2] J. W. Dunlap. Comparable tests and reliability. Journal of Educational Psychology. 24:442-453. 1933. p. 447.

From practical considerations it is usually impossible to formally evaluate each of the important objectives. Furthermore, whenever such evaluation is possible, it is necessary to determine the degree to which general objectives must be broken down into specific objectives before evaluation should be undertaken. Perhaps the most acceptable solution proposed is that objectives should be broken down until pupil reaction becomes reasonably homogeneous.

It is the purpose of this study to develop and demonstrate the use of methodological techniques which will assist in meeting the need for testing the homogeneity of behavior characteristics in psychometrics and educational evaluation.

In the foregoing discussion the problem of homogeneity resolves itself into whether or not individuals react in a significantly different manner to items between or among areas of a test than they do to items within areas. The first statistical technique which will be developed will be a test of significance based upon the null hypothesis: there is no significant difference in the manner in which individuals react to test items between or among areas of a test than to test items within areas. Since the technique involves a test of significance, acceptance or rejection of this null hypothesis is possible.

The second statistical technique to be developed and demonstrated, a proposed correction for attenuation of correlation coefficients, represents a positive indication of homogeneity rather than a test of significance. Since several formulas for correcting correlation coefficients have been previously proposed, the application of existing formulas will be demonstrated for comparison with the results obtained from the application of the proposed correction formula which permits wider application than existing formulas.

A. Review of Literature

The desirability of constructing tests which are relatively homogeneous has been recognized for some time. In 1923, Kelley formulated three rules which would assist test constructors in making comparable forms of tests.

> The following rules for the construction of two comparable tests may be laid down: (1) sufficient fore-exercise should be provided to establish an attitude or set, thus lessening the likelihood of the second test being different from the first, due to a new level of familiarity with the mechanical features, etc., (2) the elements of the first test should be as similar in difficulty and type to those in the second, pair by pair, as possible; but, (3) should not be so identical in word or form as to commonly lead to a memory transfer or correlation between errors.[1]

The foregoing quotation represents a recognition of the desirability of homogeneity for the interpretation of reliability coefficients whereby similarity of the forms of a test as to content is determined by inspection and rests with the judgment of the examiner. It does not represent a method of statistical analysis. However, the remarks are directed toward homogeneity within test forms and not toward homogeneity among areas of a test or tests measuring different behavior characteristics.

---

[1] T. L. Kelley. Statistical method. New York. The MacMillan Company. 1923. p. 203.

Again with regard to reliability, Dunlap[1] has proposed that comparability of test forms within a total test can be determined by dividing the test into fourths or constructing four forms of the total test. These four forms or the test fourths are administered to a group of subjects and the six intercorrelations computed. If the tetrads resulting from the intercorrelations all equal zero within the sampling error, the forms are said to be comparable. Thus, Dunlap concludes, "The tetrad technique offers a means of determining whether or not the split fourths or four forms of a test measure the same thing."[2]

Whereas earlier recognition of the desirability for constructing homogeneous tests emphasized the standpoint of reliability, the "technic of homogeneous tests" proposed by Loevinger[3] was developed as an alternative to the concept of reliability. The original concept of homogeneity as defined by Loevinger was restricted to tests of ability, but the technique was later generalized to include other types of tests.[4]

---

[1]J. W. Dunlap. Comparable tests and reliability. The Journal of Educational Psychology. 24:442-453. 1933.

[2]Ibid. p. 453.

[3]Jane Loevinger. A systematic approach to the construction and evaluation of tests of ability. Psychological Monographs. Vol. 61, no. 4. 1947.

[4]Jane Loevinger. The technic of homogeneous tests compared with some aspects of "Scale Analysis" and factor analysis. Psychological Bulletin. 45:507-529. 1948.

Loevinger has distinguished between perfectly homogeneous and perfectly heterogeneous tests in terms of probability:

> In a perfectly homogeneous test, when the items are arranged in the order of increasing difficulty, if any item is known to be passed, the probability is unity of passing all previous items. In a perfectly heterogeneous test, the probability of an individual passing a given item A is the same whether or not he is known already to have passed another item B.[1]

As an indication of the degree of homogeneity possessed by a test, the coefficient of homogeneity[2], H_t was derived.

$$H_t = \frac{V_x - V_{het}}{V_{hom} - V_{het}}$$

where V_x is the obtained variance of a test, V_{het} is the variance of a perfectly heterogeneous test with the same distribution of item difficulties. In computational form, the formula was expressed[3]

$$\frac{N(\Sigma Xk^2 - \Sigma Xk) + \Sigma N_i^2 - (\Sigma Xk)^2}{2N(\Sigma iN_i - \Sigma X_k) + \Sigma N_i^2 - (\Sigma Xk)^2}$$

where X refers to raw scores, N_i refers to the number passing the i<u>th</u> item, when the items are ordered according to decreasing

---

[1] Jane Loevinger. A systematic approach to the construction and evaluation of tests of ability. p. 29.

[2] <u>Ibid</u>., formula (7), p. 4.

[3] <u>Ibid</u>., formula (11), p. 4.

number passing, the subscript k means summation for all N individuals, and the subscript i means summation for all m items.

Loevinger suggested that the proposed estimate of the coefficient of homogeneity be restricted to use with samples of over 100 cases, since the magnitude of the expected sampling error under various conditions had not been determined. The sampling properties of the estimate of homogeneity also had not been ascertained at the time of her study.

The foregoing procedures represent direct methods of determining the homogeneity of a test. Although methods of item analysis are primarily concerned with test validity, several of these methods have indirect influence upon the homogeneity of the final test. Since such methods are only indirect measures of homogeneity, however, they will not be reviewed in detail. Methods of item analysis influencing homogeneity include "the precision method" illustrated by the work of Paterson[1], and Uhrbrock and Richardson[2], "item homogeneity with total test" proposed by Loevinger[3], "the biserial r method" and "the critical ratio method"

---

[1]D. G. Paterson. Preparation and use of new-type examinations. New York, World Book Company. 1935.

[2]R. S. Uhrbrock and M. W. Richardson. Item analysis. Personnel Journal. 12:141-154. 1933.

[3]Jane Loevinger. A systematic approach to the construction and evaluation of tests of ability. Psychological Monographs. Vol. 61. No. 4. 1947.

described by Guilford[1], "the Vincent overlapping method" proposed by Vincent[2], "the McCall method" and "the modified Vincent method" described by Barthelmess[3], "Long's overlapping method" and "Long's weighted overlapping method" proposed by Long[4], "the method of successive pools" attributed to Toops[5], "the method of successive residuals" proposed by Horst[6], "the summation method of agreement" proposed by Lentz and others[7], "the index of validity" proposed by Clark[8], "the index of discrimination" presented by

---

[1]J. P. Guilford. Psychometric methods. New York. McGraw-Hill Book Company. 1936.

[2]L. Vincent. A study of intelligence test elements. Teachers College, Columbia University. Contributions to Education no. 152. 1924.

[3]H. M. Barthelmess. The validity of intelligence test elements. Teachers College, Columbia University. Contributions to Education no. 505. 1931.

[4]J. A. Long. Improved overlapping methods for determining validities of test items. Journal of Experimental Education. 2:264-267. 1934.

[5]P. Horst. Item analysis by the method of successive residuals. Journal of Experimental Education. 2:254-263. 1934.

[6]P. Horst. Increasing the efficiency of selection tests. Personnel Journal. 12:254-259. 1934.

[7]T. F. Lentz, B. Hirshstein, and F. H. Finch. Evaluation of methods of evaluating test items. Journal of Educational Psychology. 23:344-350. 1932.

[8]E. L. Clark. A method of evaluating the units of a test. Journal of Educational Psychology. 19:263-265. 1928.

Cook[1], the table "the correlation of item with total score according to proportions of success" proposed by Flanagan[2], and "the equal appearing intervals method of attitude scale construction" developed by Thurstone[3].

In contrast to measures of homogeneity as properties of a test as given, Guttman[4] has proposed the concept of the determination of the "scalability" of a "universe of attributes" from which the test items constitute a sample. The items are classified according to content, and decision as to whether a universe of content constitutes a scale is made according to the "coefficient of reproducibility". This coefficient is based upon the assumption that if all items were arranged in order of difficulty, each score would represent a given "scale pattern". The coefficient of reproducibility is the percentage of responses which are reproducible from the individual's score. A "coefficient of reproducibility" of 85 to 90 per cent has been suggested as a criterion for judging "scalability". The

---

[1]W. W. Cook. The measurement of general spelling ability involving controlled comparisons between techniques. University of Iowa Studies in Education. Vol. 6: no. 6. pp. 38-59. 1932.

[2]J. C. Flanagan. General considerations in the selection of test items, etc. Journal of Educational Psychology. 30:674-680. 1939.

[3]L. L. Thurstone and E. J. Chave. The measurement of attitude. Chicago. The University of Chicago Press. 1929.

[4]Louis Guttman. A basis for scaling qualitative data. American Sociological Review. 9:139-150. 1944.

hypothesis that the universe of attributes is scalable, however, is not formulated in terms of probability.

Methods of tabulation of test data preparatory to scale analysis have included the "tabulation technique" proposed by Goodenough[1], the scalogram board, developed by Guttman[2], and the "Cornell technique", developed by Guttman.

Among the positive measures of the extent to which two or more tests or areas within a test elicit homogeneous behavior from subjects, correction for attenuation is appropriate and has been used for some time.

Since test scores are subject to errors of observation, each score deviates from the true scores for a particular subject to whom the test is administered. Although the test scores are correlated to determine the relationships between two distributions of scores, the errors of observation within each distribution may be uncorrelated, thus lowering the apparent degree of correlation. The reduction of the apparent degree of correlation comprises part of the factors which the two distributions of scores do not have in common. If the true correlation between two distributions of test scores is desired, it is necessary to make some correction for the attenuation in the raw coefficient.

---

[1]W. H. Goodenough. A technique for scale analysis. Educational and Psychological Measurement. 4:179-190. 1944.

[2]Louis Guttman. The Cornell technique for scale and intensity analysis. Educational and Psychological Measurement. 7:247-279. 1947.

As early as 1904 Spearman[1] proposed a formula for correction for attenuation. The proof of the formula was published by the same author in 1907[2] and by Brown[3] in 1909. The proof along a slightly different line was published again by Spearman[4] in 1910.

$$r_{xy}^4 = \frac{r_{x_1y_1} \cdot r_{x_2y_2} \cdot r_{x_1y_2} \cdot r_{x_2y_1}}{(r_{x_1x_2})^2 (r_{y_1y_2})^2}$$

Where r_{xy} = the true correlation between two variables x and y

x_1 = an observation of x parallel to x_2

x_2 = an observation of x parallel to x_1

y_1 = an observation of y parallel to y_2

y_2 = an observation of y parallel to y_1

---

[1]C. Spearman. The proof and measurement of association between two things. American Journal of Psychology. 15:72-101. 1904.

[2]C. Spearman. Demonstration of formulae for true measurement of correlation. American Journal of Psychology. 18:161-169. 1907.

[3]William Brown. Some experimental results in correlation. Sixth International Congress of Psychology, Geneva, 1909. Proceedings p. 571-576. 1910.

[4]C. Spearman. Correlation calculated from faulty data. British Journal of Psychology. 3:271-295. 1910.

$r_{x_1y_1}$ = the correlation coefficient between x_1 and y_1

$r_{x_2y_2}$ = the correlation coefficient between x_2 and y_2

$r_{x_1y_2}$ = the correlation coefficient between x_1 and y_2

$r_{x_2y_1}$ = the correlation coefficient between x_2 and y_1

$r_{x_1x_2}$ = the correlation coefficient between x_1 and x_2

$r_{y_1y_2}$ = the correlation coefficient between y_1 and y_2

In the development of the formula it was assumed that of the six quantities x, y, σ_x, σ_y, errors of observation$_1$, and errors of observation$_2$, x and y alone are correlated.

Cureton and Dunlap[1] have shown, however, that Spearman's original formula using the geometric mean of the intercorrelations as the numerator is an underestimate of the true correlation. They indicate that the same formula, altered so as to use the arithmetic mean of the intercorrelations in the numerator,

$$r_{xy} = \frac{r_{x_1y_1} + r_{x_2y_2} + r_{x_1y_2} + r_{x_2y_1}}{4\sqrt{(r_{x_1x_2})(r_{y_1y_2})}}$$

---

[1]E. E. Cureton and J. W. Dunlap. Spearman's correction for attenuation and its probable error. American Journal of Psychology. 42:235-245. 1930. Formula [5].

although still an underestimate, is nearer the true r.

> It should be noted that the factor causing a systematic error in the estimation of r_{xy} in the case of formula [5] [Proposed formula with arithmetic mean in the numerator] is inequality of standard deviations, while formula [6] [Spearman's formula with geometric mean in the numerator] it is inequality of product-moments. In each instance the tendency is toward underestimation, while both formulas have the same denominator, which is not subject to any systematic error. The numerator of [5] is the arithmetic mean of the four intercorrelations while that of [6] is their geometric mean. Hence [5] will always yield the less underestimated value.[1]

It has been shown by Kelly[2] that it is not necessary to compute $r_{x_1y_1}$, $r_{x_2y_2}$, $r_{x_1y_2}$ and $r_{x_2y_1}$ if the value of $r_{(x_1 + y_1)(x_2 + y_2)}$ is known; i.e., the value of the correlation between the sums of the scores for the forms of each test.

Cureton and Dunlap[3] have given proof for the use of $r_{(x_1 + y_1)(x_2 + y_2)}$ into the correction for attenuation. The

---

[1] Ibid., p. 237.

[2] T. L. Kelly. Statistical method. The MacMillan Company. New York. 1923. pp. 209-210.

[3] E. E. Cureton and J. W. Dunlap. Op. cit. p. 238. Formula [8].

formula is

$$r_{xy} = \frac{r_{(x_1 + y_1)(x_2 + y_2)}}{\sqrt{\frac{2r_{x_1x_2}}{1 + r_{x_1x_2}}}\sqrt{\frac{2r_{y_1y_2}}{1 + r_{y_1y_2}}}}$$

This formula is implied in Kelly's derivation of the substitution of $r_{(x_1 + y_1)(x_2 + y_2)}$ for the average of $r_{x_1y_1}$, $r_{x_2y_2}$, $r_{x_1y_2}$ and $r_{x_2y_2}$, and was published in this form without formal proof by Hull.[1]

Peters and VanVoorhis[2] have shown the development of a formula for correction for attenuation in the following form:

$$r_{xy} = \frac{\bar{r}_{xy}}{\sqrt{r_{x_{12}} r_{y_{12}}}}$$

in which $\bar{r}_{xy}$ is the average intercorrelation of any number of samples of measures of the x and y functions. These authors, however, did not suggest the method which should be used in obtaining $\bar{r}_{xy}$.

---

[1] C. L. Hull. Aptitude testing. World Book Company, New York. 1928. p. 243. Formula [5].

[2] Charles C. Peters and Walter R. VanVoorhis. Statistical procedures and their mathematical bases. McGraw-Hill Book Company, New York. 1940. p. 203.

II. DEVELOPMENT OF PROPOSED FORMULAS

A. F-test for Departure from Homogeneity

For the development of the proposed formula for testing the significance of departure from homogeneity, let it be supposed that a test is given to N subjects. There are "k" areas included in the test and two forms are made for each area. The forms within each area represent two equivalent groups of items and are determined by selecting items from the area at random, thus yielding the two forms or subtests. Let it be assumed that the items within each area have been assembled in random order; the two forms are made by choosing odd- and even-numbered items.

An analysis of variance table may then be set up as in Table 1, with the needed sums of squares and mean squares indicated.

The test of significance desired consists of noting whether the fluctuations of scores from subject to subject are significantly greater between areas than within areas, i.e.,

$$F = \frac{\text{Mean Square for Subject by Area}}{\text{Mean Square for Subject by Form}}$$

For convenience in the development of the formula, let all scores within each form be reduced to standard scores,

$$Z = \frac{X - m}{\sigma},$$

thus forcing equal means and variances for each form in each area.

Table 1

Analysis of Variance - Two Forms and "k" Areas

Source of Variation	Degrees of Freedom	Sum of Squares	Mean Square
Subject	N - 1	needed	
Area	k - 1	needed	
Form	1	needed	
Subject x Area	(N - 1)(k - 1)	needed	needed
Subject x Form	N - 1	needed	needed
Area x Form	k - 1		
Subject x Area x Form	(k - 1)(N - 1)		
Total	2Nk - 1		

The necessary sums of squares to test the hypothesis are then computed. The usual correction term can be disregarded since it is zero for standard scores. The sum of squares for Subject

$$= \frac{\Sigma[z_1 + z_2 \ldots + z_{2k}]^2}{2k}$$

Let subscripts of an odd number and the next even number be assigned to the same area, the sum of squares for interaction for Subject by Area

$$= \frac{\Sigma[z_1 + z_2]^2 + \Sigma[z_3 + z_4] + \ldots + \Sigma[z_{2k-1} + z_2]^2}{2} - \frac{\Sigma[z_1 + z_2 + \ldots + z_{2k}]^2}{2k}$$

$$= \frac{k\Sigma[z_1 + z_2]^2 + k\Sigma[z_3 + z_4] + \ldots + k\Sigma[z_{2k-1} + z_2]^2 - \Sigma[z_1 + z_2 + \ldots + z_{2k}]^2}{2k}$$

Let subscript be assigned

A = interarea subtests

W = intra-area subtests

T = all subtests.

Then the foregoing expression becomes

$$= \frac{k\Sigma z_T^2 + 2k\Sigma z_A z_A - 2k\Sigma z_A z_A - \Sigma z_T^2 - 2\Sigma z_T z_T}{2k}$$

or, since

$$\Sigma z_T z_T = \Sigma z_A z_A + \Sigma z_W z_W$$

the expression may be written

$$= \frac{(k-1)\ \Sigma z_T^2 + 2(k-1)\ \Sigma z_W z_W - 2\Sigma z_A z_A}{2\ k}$$

But

$$\Sigma z^2 = N$$

for each subtest. Since the number of subtests is 2k, therefore,

$$\Sigma z_T^2 = 2Nk.$$

Also

$$\Sigma zz = Nr,$$

Therefore, the expression for the sum of squares

$$= \frac{Nk(k-1) + N(k-1)\ \Sigma r_W - N\Sigma r_A}{k\ (k-1)(N-1)}$$

$$= \frac{N}{N-1}\left[1 + \frac{\Sigma r_W}{k} - \frac{\Sigma r_A}{k\ (k-1)}\right]$$

For the present let it be assumed that the average r $(\bar{r})$ can be obtained by the arithmetic mean; then

$$\bar{r}_A = \frac{\Sigma r_A}{2k\ (k-1)} \quad \text{(since there are 2k (k-1) terms)}$$

and

$$\bar{r}_W = \frac{\Sigma r_W}{k} \quad \text{(since there are k terms).}$$

The mean square for Subject by Area, representing the numerator for the F-test is then

$$= \frac{N}{N-1}\left[1 + \bar{r}_W - 2\bar{r}_A\right]$$

with (N-1)(k-1) degrees of freedom.

For the F-test denominator, the mean square interaction of Subject by Form is needed. The sum of squares for Subject by Form may be determined in many ways, i.e., all possible combinations of items to develop subtests or forms within each area. In the case the odd items and the even items are taken, the interaction

$$= \frac{\Sigma[z_1 + z_3 + \ldots + z_{2k-1}]^2 + \Sigma[z_2 + z_4 + \ldots + z_{2k}]^2}{k} - \frac{\Sigma[z_1 + z_2 + \ldots + z_{2k}]^2}{2k}$$

$$= \frac{2[\Sigma z_1^2 + \Sigma z_2^2 + \ldots + \Sigma z_{2k}^2] + 4[\Sigma zz(\text{all odd}) + \Sigma zz(\text{all even})]}{2k}$$

$$- \frac{[\Sigma z_1^2 + \Sigma z_2^2 + \ldots + \Sigma z_{2k}^2] - 2[\Sigma zz(\text{for all } 2k^2-k \text{ terms})]}{2k}$$

$$= \frac{\Sigma[z_1^2 + \Sigma z_2^2 + \ldots + \Sigma z_{2k}^2] + 4[\Sigma zz(\text{all odd}) + \Sigma zz(\text{all even})]}{2k}$$

$$- \frac{2[\Sigma zz(\text{for all } 2k^2-k \text{ terms})]}{2k}$$

When a similar sum of squares has been written for each of the combinations from which form total may accrue, and an average found from them, all cross products will vanish on summation except the intra-area ones, $\Sigma z_1 z_2$, $\Sigma z_3 z_4$, etc., to $\Sigma z_{2k-1} z_{2k}$.

The average sum of squares for all possible combinations of interaction of Subject by Form

$$= \frac{[\Sigma z_1^2 + \Sigma z_2^2 + \ldots + \Sigma z_{2k}^2] - 2[\Sigma z_1 z_2 + \Sigma z_3 z_4 + \ldots + \Sigma z_{2k-1} z_{2k}]}{2k}$$

or with previous notation

$$= \frac{\Sigma z_T^2 - 2\Sigma z_W z_W}{2k} .$$

But

$$\Sigma z_T^2 = 2Nk$$

and

$$2\Sigma z_W z_W = 2N\Sigma r_W .$$

Substituting, the expression becomes

$$= \frac{2kN - 2N\Sigma r_W}{2k}$$

$$= N[1-\bar{r}_W].$$

The mean square for Subject by Form, representing the denominator of the F-test is

$$\frac{N}{N-1}\ [1-\bar{r}_W]$$

with N-1 degrees of freedom.

Significance may be tested by noting the ratio

$$F = \frac{\frac{N}{N-1}\ [1+\bar{r}_W - 2\bar{r}_A]}{\frac{N}{N-1}\ [1 - \bar{r}_W]}$$

$$= \frac{1 + \bar{r}_W - 2\ \bar{r}_A}{1 - \bar{r}_W}$$

In the development of this formula, it was assumed that it was satisfactory to average correlations by computing arithmetic means. Probably a more satisfactory method[1] is that of computing the mean of the function

$$1/2\ \log_e\ \frac{1 + r}{1 - r}\ .$$

Although the development of the formula utilized standard scores, their use is unnecessary for solving the final formula for testing departure from homogeneity. The F-value obtained may

---

[1]Paul R. Rider. Modern statistical methods. New York, John Wiley and Sons. 1939. p. 106.

be compared with those given in Snedecor's table[1] with (N-1)(k-1) and (N-1) degrees of freedom. If the F-value for departure from homogeneity is not significant, homogeneity has not been disproved. If, however, it is significant, the hypothesis of homogeneity is rejected.

It should be noted that the technique here proposed does not guarantee homogeneity within a given area, but indicates the relative greater lack of homogeneity between or among areas than within areas.

---

[1]George W. Snedecor. Statistical methods. 4th ed. Ames, Iowa. Iowa State College Press. 1946. p. 284.

B. Correction for Attenuation

The foregoing formula, F-test for the significance of departure from homogeneity, involves testing the null hypothesis. The second statistical measure to be developed, a formula for correction for attenuation, is proposed to be used as a positive indication of homogeneity rather than a test of significance.

The correction for attenuation yields an estimate of the coefficient of correlation which would prevail if the forms of the test were of infinite length, i.e., all intra-form coefficients of correlation equal to unity. Such a situation is mathematically possible, but scarcely conceivable in actual practice.

For the development of the following formula the same notation as was employed in the previous development was used. In the analysis of regression of the inter- and intra-area relationships between test scores, the sum of squares for residuals which cannot be accounted for by the reliability coefficients is equal to the difference between the sum of squares for interarea regression and the sum of squares for intra-area regression. Symbolically,

$$(1 - r_A^2)\,\Sigma y^2 - (1 - r_W^2)\,\Sigma y^2 = (1 - r_{OO}^2)\,\Sigma y^2$$

where r_{OO} is the statistical measure here desired.

Solving for r_{oo}, the formula becomes

$$(1 - r_A^2) - (1 - r_W^2) = (1 - r_{oo}^2)$$

$$r_A^2 - r_W^2 = 1 - r_{oo}^2$$

$$r_{oo}^2 = 1 - r_W^2 + r_A^2$$

$$r_{oo} = \sqrt{1 - r_W^2 + r_A^2}$$

III. EMPIRICAL JUSTIFICATION OF PROPOSED FORMULAS

A. Correlation Unity

The coefficient of correlation corrected for attenuation is the estimated correlation between true scores on two tests. If two tests are measures of the same behavior characteristic, then the correlation between them will differ from unity simply because of variable errors. A satisfactory formula for correcting a coefficient of correlation for attenuation should eliminate the effect of such errors. When variable errors are introduced into a set of numerical series whose correlation is reduced from unity only because of the variable errors which were introduced, then the degree to which the correction formula eliminates the effects of the variable errors can be observed.

To empirically justify the use of the proposed formula for correction for attenuation, and to observe in the same manner the corrections for attenuation which have been developed by Spearman, Kelly-Hull, Cureton-Dunlap, and the formula implied by Peters and VanVoorhis, two hypothetical distributions of 20 values with perfect relationship between them were designated. These hypothetical distributions are shown in Table 2 in the columns entitled "Basic$_1$" and "Basic$_2$". Since these distributions are identical, the correlation between them is unity.

To approximate the conditions of test construction in which reliability is obtained by the construction of two forms of a test by choosing odd and even-numbered items, forms 1, I, 2 and II, were developed by allowing "Basic$_1$" and "Basic$_2$" to fluctuate by adding 0 or 1 at random to each of the 20 basic values with the use of a table of random numbers. Forms 1 and I may then be interpreted as representing the split halves or subtests of a measuring instrument, form 1 being the scores obtained on the odd-numbered items and form I being the scores obtained on the even-numbered items for 20 subjects. The same interpretation may be made for forms 2 and II. The columns entitled "Total" represent the sum of the two forms of each test.

The introduction of variable errors into the originally perfectly-correlated basic distributions results in lowering the obtained correlations. Satisfactory formulas for correction for attenuation should yield estimates varying from unity only because of chance combinations. Since there are only 20 hypothetical subjects involved, the assignment of 0 and 1 at random to their scores might be expected to yield unusual estimates simply because of chance combinations among the four forms. The foregoing procedure was therefore repeated ten times to provide ten sets of distributions, each of which was made up of forms 1, I, 2 and II which represents a wider range of conditions than

would exist if fewer demonstrations had been used. The use of ten demonstrations also allowed an overall or combination of 800 values to be observed, thus lessening the probability of getting results which might have been greatly influenced by chance combination.

From the four forms constructed with the use of the basic distributions it was possible to compute four inter- and two intra-area correlations for each of the ten demonstrations and for the combined or total demonstration.

Table 2

Illustration of Demonstration Construction, Correlation Unity

Basic 1	Form 1	Form I	Total (1+I)	Basic 2	Form 2	Form II	Total (2+II)
22	23	23	46	22	23	22	45
23	23	24	47	23	23	23	46
23	24	24	48	23	24	24	48
24	24	24	48	24	25	25	50
24	25	24	49	24	25	25	50
24	25	24	49	24	24	24	48
25	26	26	52	25	25	25	50
25	26	25	51	25	26	26	52
25	26	25	51	25	25	25	50
25	26	25	51	25	25	25	50
26	26	26	52	26	26	26	52
26	27	26	53	26	27	27	54
26	26	26	52	26	27	27	53
26	26	27	53	26	27	27	53
27	27	28	55	27	27	27	55
27	27	27	54	27	28	28	55
27	27	27	54	27	27	27	55
28	29	29	58	28	28	28	57
28	29	28	57	28	29	29	57
29	29	30	59	29	29	29	58

Table 3

Inter- and Intra-Area Correlations for Ten Demonstrations, Correlation Unity

Demonstration	Correlations					
	1 I	1 2	1 II	I 2	I II	2 II
1	.919	.941	.929	.930	.905	.920
2	.924	.943	.954	.917	.925	.942
3	.935	.899	.921	.913	.920	.940
4	.947	.919	.924	.927	.915	.949
5	.930	.946	.956	.953	.945	.923
6	.953	.934	.938	.931	.922	.928
7	.933	.896	.896	.897	.898	.966
8	.909	.929	.968	.966	.937	.955
9	.955	.938	.912	.965	.933	.933
10	.904	.899	.943	.925	.924	.941
Total	.926	.920	.929	.930	.919	.935

Table 4

Spearman-Brown Estimates for Intra-Area Correlations and Total Form Correlations for Ten Demonstrations, Correlation Unity

Demonstration	Spearman-Brown Estimate $r_{1\ I}$	Spearman Brown Estimate $r_{2\ II}$	Total Form Correlation $r_{(1+I)(2+II)}$
1	.958	.958	.964
2	.961	.970	.968
3	.966	.969	.946
4	.973	.974	.946
5	.964	.960	.986
6	.976	.963	.960
7	.965	.983	.919
8	.952	.977	.983
9	.977	.965	.963
10	.950	.970	.961
Total	.962	.966	.968

Table 5

Average Inter- and Intra-Area Correlations for Ten Demonstrations, Correlation Unity

Demonstration	Average Interarea Correlation ($\bar{r}_A$)	Average Intra-Area Correlation ($\bar{r}_W$)
1	.902	.927
2	.936	.934
3	.914	.938
4	.922	.948
5	.950	.927
6	.931	.942
7	.897	.952
8	.953	.936
9	.940	.945
10	.925	.931
Total	.931	.925

The intercorrelations for the ten demonstrations and the total demonstration are shown in Table 3. With the exception of the seventh demonstration in which the interarea correlations are all relatively low, the correlation coefficients are distributed in no discernible pattern.

In Table 4 are shown the Spearman-Brown estimates for the correlations between forms 1 - I and 2 - II after having been stepped up by the formula $\frac{2r}{1+r}$. The total form correlations, (1+I)(2+II), are also shown in this table. The Spearman-Brown estimates, and the total form correlations are used in the solution of the Kelly-Hull formula for correction for attenuation.

In Table 5 are shown the average inter- and intra-area correlations for the ten demonstrations and for the total demonstration. The average interarea correlations, $\bar{r}_A$, represent the average of the four interarea correlations, $r_{1\,2}$, $r_{1\,II}$, $r_{I\,2}$ and $r_{I\,II}$. The average intra-area correlations, $\bar{r}_W$, represent the average of the two intercorrelations, $r_{1\,I}$ and $r_{2\,II}$. Both averages, $\bar{r}_A$ and $\bar{r}_W$, have been obtained by use of the function

$$1/2 \log_e \frac{1+r}{1-r}.$$

The average inter- and intra-area correlations shown in Table 5 are utilized in the solution of the proposed formula for correction for attenuation.

From the data shown in Tables 3, 4 and 5, it was possible to compute the corrections for attenuation according to the five formulas previously discussed. The various estimates of the true correlation in the distributions for each of the ten demonstrations and for the combined demonstrations are shown in Table 6.

In the first column of Table 6, entitled Correction for Attenuation, Kelly-Hull, are shown the estimates of the true correlation in the ten demonstrations and the combined demonstration computed with the use of the formula

$$r_{oo} = \frac{r_{(1+I)(2+II)}}{\sqrt{\frac{2r_{1\ I}}{1+r_{1\ I}}}\ \sqrt{\frac{2r_{2+II}}{1+r_{2+II}}}}$$

In the second column of Table 6 are shown the correlations corrected for attenuation by the use of Spearman's formula

$$r^4_{oo} = \frac{r_{1\ 2} \cdot r_{1\ II} \cdot r_{I\ 2} \cdot r_{I\ II}}{(r_{1\ I})^2\ (r_{2\ II})^2}$$

The corrected correlations in the third column have been obtained by using Cureton and Dunlap's formula

$$r_{oo} = \frac{r_{1\,2} + r_{1\,II} + r_{I\,2} + r_{I\,II}}{4\sqrt{(r_{1\,I})(r_{2\,II})}}$$

In the fourth column are shown the correlations corrected for attenuation with the use of the formula

$$r_{oo} = \frac{\bar{r}_W}{\sqrt{(r_{1\,I})(r_{2\,II})}}$$

which has been suggested by Peters and Van Voorhis. Although no method for averaging the r's in the numerator was proposed, the function

$$1/2 \log_e \frac{1 + r}{1 - r}$$

was used here.

In the fifth column the correlations have been corrected for attenuation by using the formula here proposed

$$r_{oo} = \sqrt{1 - (\bar{r}_W)^2 + (\bar{r}_A)^2}$$

In addition to the correlations corrected for attenuation shown in Table 6, the ten demonstrations were treated by the use of the F-test for significance of departure from homogeneity. The proposed formula is

$$F = \frac{1 + \bar{r}_W - 2\bar{r}_A}{1 - \bar{r}_W}$$

Table 6

Corrections for Attenuation and F-values for Ten Demonstrations, Correlation Unity

Demonstration	Correction for Attenuation					F-Value
	Kelly-Hull	Spearman	Cureton-Dunlap	Peters-VanVoorhis	Proposed	
1	1.006	1.007	1.007	1.008	1.023	.49
2	1.003	1.017	1.074	1.075	1.001	1.38
3	.978	.974	.974	1.001	.978	.44
4	.972	.985	.972	1.000	.975	.33
5	1.025	1.025	1.018	1.001	1.021	1.92
6	.990	.985	.990	1.002	.990	.68
7	.944	.945	.945	.997	.948	.07
8	1.019	1.020	1.020	1.005	1.016	1.72
9	.992	.999	.993	1.001	.996	.83
10	1.001	1.000	1.000	.991	.994	.84
Total	1.004	.994	.994	.994	1.005	1.03

In none of the demonstrations was the F-value for departure from homogeneity significant. Fluctuations among the sizes of the ten F-values in Table 6 are small. The F-value for the combined demonstrations closely approximates the variance ratio which would be expected from duplicate distributions and homogeneity has not been disproven.

Inspection of the correlations corrected for attenuation in Table 6 indicates that the five formulas are all relatively satisfactory for removing the effect of errors of observation under the circumstances imposed in the demonstrations. Differences among the estimated true correlations are so slight in most cases as to be attributable to systematic errors of rounding numbers in the process of computation.

From the ten foregoing demonstrations it was concluded that the five formulas for correction for attenuation were equally satisfactory for removing the effect of variable errors of observation when similarity of means and standard deviations in the four test forms was imposed.

B. Correlation Less Than Unity

In the foregoing demonstrations of the formulas for correction for attenuation the condition of approximately equal variances in the forms was imposed. Further, the hypothetical reliabilities of the two forms throughout the demonstrations were all relatively high. In actual practice, this condition will seldom prevail. It was considered desirable to develop demonstrations in which the estimates of the true correlation might be observed and in which differences between reliabilities would be imposed.

To approximate test construction conditions of measurement where reliabilities of the two forms are not equal, the two basic distributions used in the former demonstrations were altered in a slightly different manner than that used for constructing the demonstration in which the true correlation was unity. To develop two sets of odd-even distributions which would yield unequal reliabilities, $Basic_1$ was made to fluctuate 0 or 1 at random and $Basic_2$ was made to fluctuate 0 to 9 at random. An illustration of this procedure is shown in Table 7.

Forcing unequal reliabilities lowers the true correlation from unity and the extent to which this is done is indefinite. In the following ten demonstrations the true correlation is somewhat

Table 7

Illustration of Demonstration Construction, Correlation Less Than Unity

Basic 1	Form 1	Form I	Total (1+I)	Basic 2	Form 2	Form II	Total (2+II)
22	23	23	46	22	26	24	50
23	23	24	47	23	32	27	59
23	24	24	48	23	28	32	60
24	24	24	48	24	29	30	59
24	25	24	48	24	29	30	59
24	25	24	49	24	28	30	59
25	26	26	52	25	33	33	66
25	26	25	51	25	28	29	57
25	26	25	51	25	30	28	58
25	26	25	51	25	27	27	54
26	26	26	52	26	27	31	58
26	27	26	53	26	28	35	63
26	26	26	52	26	29	27	56
26	26	27	53	26	32	30	62
27	27	28	55	27	34	34	68
27	27	27	54	27	34	27	61
27	27	27	54	27	28	35	63
28	29	29	58	28	36	31	67
28	29	29	58	28	37	34	71
29	29	30	59	29	38	35	73

less than unity, though high. Some variation in the size of the estimates of the true correlation would be expected since only 20 values were used in constructing each form. However, the method of procedure in constructing each demonstration was the same.

The four interarea and two intra-area correlations for each of the ten demonstrations and the combined or total demonstration are shown in Table 8. All the intra-area correlations for the forms developed from $Basic_1$, $r_{1\ I}$, are higher than the intra-area correlations obtained from $Basic_2$, $r_{2\ II}$. In the third demonstration $r_{2\ II}$ is negative, which prohibits the application of all but the proposed correction for attenuation formula in this demonstration.

The Spearman-Brown estimates for each form and for the total form correlations are shown in Table 9. The negative correlation in the third demonstration made the Spearman-Brown estimates inappropriate in this instance.

The average inter- and intra-area correlations obtained by the function $1/2 \log_e \frac{1 + r}{1 - r}$ are shown in Table 10.

The results of solving the various formulas for correction for attenuation are shown in Table 11 with the F-values for significance of departure from homogeneity for the ten demonstrations and for the combined demonstrations.

Table 8

Inter- and Intra-Area Correlations for Ten Demonstrations, Correlation Less Than Unity

Demonstration	Correlations					
	$r_{1\ II}$	$r_{1\ 2}$	$r_{1\ II}$	$r_{I\ 2}$	$r_{I\ II}$	$r_{2\ II}$
1	.919	.672	.594	.805	.629	.475
2	.924	.547	.278	.659	.511	.497
3	.935	.450	.247	.508	.283	-.114
4	.944	.509	.556	.488	.542	.394
5	.930	.481	.377	.416	.351	.491
6	.953	.395	.547	.392	.434	.102
7	.933	.769	.590	.794	.500	.575
8	.909	.573	.497	.575	.591	.604
9	.955	.637	.528	.619	.477	.648
10	.904	.472	.493	.466	.588	.351
Total	.926	.537	.469	.561	.484	.409

Table 9

Spearman-Brown Estimates for Intra-Area Correlations and Total Form Correlations for Ten Demonstrations, Correlations Less Than Unity

Demonstration	Spearman-Brown Estimate $r_{1\ I}$	Spearman-Brown Estimate $r_{2\ II}$	Total Score Correlation $r_{(1+I)(2+II)}$
1	.958	.644	.807
2	.960	.664	.637
3	-	-	.588
4	.971	.565	.637
5	.964	.659	.545
6	.976	.185	.600
7	.965	.730	.764
8	.952	.753	.638
9	.972	.786	.634
10	.950	.520	.632
Total	.962	.581	.630

Table 10

Average Inter- and Intra-Area Correlations for Ten Demonstrations, Correlation Less Than Unity

Demonstration	Average Interarea Correlation ($\bar{r}_A$)	Average Intra-Area Correlation ($\bar{r}_W$)
1	.685	.782
2	.511	.793
3	.377	.659
4	.524	.799
5	.410	.799
6	.444	.754
7	.682	.824
8	.560	.803
9	.569	.869
10	.507	.731
Total	.514	.775

Table 11

Corrections for Attenuation and F-values for Ten Demonstrations, Correlation Less Than Unity

Demonstration	Correction for Attenuation					F-Value
	Kelly-Hull	Spearman	Cureton-Dunlap	Peters-VanVoorhis	Proposed	
1	1.027	1.017	1.022	1.037	.926	1.89
2	.798	.702	.736	.754	.794	3.72**
3	-	-	-	-	.994	2.65*
4	.860	.858	.859	.859	.801	3.74**
5	.684	.597	.601	.607	.746	4.87**
6	1.412	1.405	1.418	1.424	.793	3.52**
7	.910	.914	.906	.931	.886	2.61*
8	.754	.753	.754	.756	.816	3.47**
9	.725	.714	.719	.723	.754	5.58**
10	.899	.892	.896	.900	.850	2.67*
Total	.843	.831	.833	.835	.815	3.32**

Required for significance 5% 2.15*
1% 3.00**

The degree of significance of the departure from homogeneity indicated by the F-value is designated by two asterisks for one per cent significance and one asterisk for five per cent significance. With 19 and 19 degrees of freedom, a value of 2.15 is required for five per cent significance and 3.00 for one per cent significance, and with 199 and 199 degrees of freedom, 1.26 and 1.39, respectively.

Of the ten demonstrations and the combined demonstration, the hypothesis of homogeneity is refuted at the five per cent level of confidence in ten demonstrations and at the one per cent level in seven. In only the first demonstration is nonhomogeneity not indicated and in this demonstration all corrections for attenuation have yielded estimates close to unity.

From inspection of the relative sizes among the F-values and the relative sizes among the correlations corrected for attenuation by use of the proposed formula, it is apparent that the relation of the demonstrations within these two measures is not exactly the same. This discrepancy results from basing the F-value on evidence available from the existing correlations, whereas the corrected correlations are estimates of the relation existing with perfectly reliable forms of a test.

The four formulas for correction for attenuation, Kelly-Hull, Spearman, Cureton-Dunlap and Peters-VanVoorhis all yielded varying estimates among the ten demonstrations. Although the practice of

reporting corrected correlations as being no higher than unity[1] has been proposed, formulas which yield estimates considerably higher than unity in some instances suggest serious consideration of the accuracy of the estimates yielded in other instances.

Of the five formulas for correction for attenuation, the proposed formula is the only one which is not rendered inappropriate by the presence of the negative correlation in the third demonstration.

The five formulas for correction for attenuation yielded estimates for the combined demonstrations which do not vary greatly. Combining the ten demonstrations has apparently had a tendency to eliminate the effect of unusual fluctuations resulting from the small number of cases in each demonstration.

It was concluded from these demonstrations that the proposed formula for correction for attenuation yielded estimates of the true correlation which were more uniform than the estimates yielded by the use of the other formulas. When greater numbers of cases were considered, however, the five formulas did not vary greatly in the size of their estimates.

---

[1]Charles C. Peters and Walter R. VanVoorhis. Statistical procedures and their mathematical bases. New York. McGraw-Hill Book Company. 1940. p. 204.

C. Unlike Basic Distributions

Under many circumstances of test construction an estimate of the true correlation between measurements of dissimilar behavior characteristics is desired. In such a situation it is unlikely that the true correlation would ever be unity. It is also unlikely that the reliabilities would be equal.

To approximate a testing situation in which the reliabilities are unequal and the true correlation is less than unity, hypothetical test forms were constructed from "Basic$_1$" and "Basic$_3$" shown in Table 12. Forms 1 and I were constructed by making "Basic$_1$" fluctuate 0 or 1 at random. "Basic$_3$" was constructed by making "Basic$_1$" fluctuate 0 to 3 at random. Forms 2 and II were constructed by making "Basic$_3$" fluctuate 0 to 9 at random.

The foregoing procedure was again repeated ten times to yield ten demonstrations using 20 cases and a combined or total demonstration of 200 cases.

The six inter- and intra-area correlations were computed for the four forms in the demonstrations. These inter- and intra-area correlations are shown in Table 13. The correlation between forms 2 and II in the fourth demonstration is negative. Whereas the intra-area correlations for 1 and I are relatively uniform throughout the ten demonstrations, considerable variation is evident

among the interarea correlations, 1 2, 1 II, I 2, I II, and the intra-area correlations, 2 II.

The Spearman-Brown estimates for the intra-area correlations and the total form correlations are shown in Table 14. The Spearman-Brown estimates were inappropriate for the fourth demonstration because of the negative value for $r_{2\ II}$.

The average inter- and intra-area correlations obtained by the use of the function

$$1/2 \log_e \frac{1 + r}{1 - r}$$

are shown in Table 15.

The data in Tables 13, 14 and 15 yielded coefficients of correlations corrected for attenuation according to the results shown in Table 16. The F-values for significance of departure from homogeneity for the demonstrations based on unlike basic distributions are also shown in Table 16.

Table 12

Illustration of Demonstration Construction, Unlike Basic Distributions

$Basic_1$	Form 1	Form I	Total (1+I)	$Basic_3$	Form 2	Form II	Total (2+II)
22	23	22	45	24	27	26	53
23	23	23	46	26	35	28	63
23	24	24	48	25	27	26	53
24	25	25	50	28	34	37	71
24	25	25	50	26	32	34	66
24	25	24	49	26	31	26	57
25	26	25	51	26	34	32	66
25	26	26	52	27	33	31	64
25	25	25	50	26	28	31	59
25	25	25	50	28	35	37	72
26	27	26	53	29	30	38	68
26	27	27	54	27	30	35	65
26	26	27	53	30	39	37	76
26	26	27	53	28	37	33	70
27	28	27	55	29	29	38	67
27	27	28	55	31	39	33	72
27	28	27	55	30	36	38	74
28	29	28	57	31	31	38	69
28	28	29	57	30	33	39	72
29	29	29	58	30	31	30	61

Table 13

Inter- and Intra-Area Correlations for Ten Demonstrations, Unlike Basic Distributions

Demonstration	Correlations					
	1 I	1 2	1 II	I 2	I II	2 II
1	.920	.131	.629	.334	.610	.492
2	.942	.591	.597	.574	.586	.620
3	.940	.667	.740	.662	.750	.428
4	.949	.298	.185	.371	.108	- .083
5	.923	.511	.324	.559	.489	.142
6	.928	.483	.542	.405	.362	.154
7	.966	.552	.546	.606	.581	.654
8	.955	.682	.514	.674	.508	.635
9	.933	.802	.792	.718	.280	.140
10	.941	.611	.662	.528	.633	.499
Total	.935	.519	.721	.522	.467	.341

Table 14

Spearman-Brown Estimates for Intra-Area Correlations and Total Form Correlations for Ten Demonstrations, Unlike Basic Distributions

Demonstration	Spearman-Brown Estimate $r_{1\ I}$	Spearman-Brown Estimate $r_{2\ II}$	Total Form Correlation $r_{(1+I)(2+II)}$
1	.958	.660	.560
2	.970	.765	.650
3	.969	.599	.845
4	-	-	.366
5	.960	.249	.745
6	.963	.267	.595
7	.983	.791	.633
8	.997	.777	.652
9	.965	.246	.644
10	.970	.666	.711
Total	.966	.509	.630

Table 15

Average Inter- and Intra-Area Correlations for Ten Demonstrations, Unlike Basic Distributions

Demonstration	Average Interarea Correlation ($\bar{r}_A$)	Average Intra-Area Correlation ($\bar{r}_W$)
1	.448	.787
2	.587	.845
3	.707	.799
4	.244	.701
5	.475	.704
6	.451	.716
7	.572	.886
8	.601	.866
9	.677	.766
10	.611	.816
Total	.567	.772

Inspection of the F-values shown in Table 16 indicates that the combined demonstrations yielded departure from homogeneity significant at the five per cent level but not significant at the one per cent level of confidence. The F-values varied more among the ten demonstrations in Table 16 than the ten demonstrations in Table 11 or 6. Of the 11 F-values in Table 16, three reached five per cent but not one per cent significance, and six others were significant at the one per cent level.

As in Table 6, the proposed formula yielded more uniform estimates of the true correlation than the other formulas for correction for attenuation. None of the estimates of the true correlation exceeded unity when the proposed formula was used for the correction. The negative correlation between 2 II in the fourth demonstration rendered the other formulas inappropriate.

Table 16

Corrections for Attenuation and F-values for
Ten Demonstrations, Unlike Basic Distributions

Demon-stration	Correction for Attenuation					F-Value
	Kelly-Hull	Spearman	Cureton-Dunlap	Peters-VanVoorhis	Proposed	
1	.704	.535	.633	.667	.762	4.18**
2	.754	.768	.768	.768	.794	4.33**
3	1.109	1.110	1.111	1.115	.928	1.92
4	-	-	-	-	.754	4.06**
5	1.524	1.275	1.300	1.312	.854	2.55*
6	1.173	1.173	1.185	1.193	.831	2.87*
7	.718	.718	.719	.720	.736	6.51**
8	.757	.756	.763	.772	.782	4.96**
9	1.322	1.654	1.793	1.873	.968	1.36
10	.885	.885	.888	.892	.841	3.23**
Total	.898	.978	.987	1.004	.852	2.80*

Required for significance 5% 2.15*
1% 3.00**

IV. APPLICATION OF PROPOSED FORMULAS

To demonstrate the application of the F-test for significance of departure from homogeneity and the proposed correction for attenuation, several situations in which the formulas are appropriate were selected. The test situations for this demonstration were chosen to represent different types of circumstances where the formulas can be applied. In choosing the test situations emphasis was placed on availability of test data and appropriateness, and only one test of each of the various types was used. The purpose of the following demonstrations of application is to illustrate the method of applying the formulas rather than to determine characteristics of the specific testing instruments themselves.

A. Aptitude

The American Council on Education Psychological Examination for college freshmen yields scores for two general areas of mental processes. Scores on the following subtests are combined to form the Quantitative score:

Arithmetic Reasoning
Number Series
Figure Analogies.

Scores on the following subtests are combined to form the Linguistic score:

Same - Opposites
Completion
Verbal Analogies.

Twenty-three minutes testing time are allowed for the Quantitative area and fifteen minutes for the Linguistic area.

Separate norms for the two sets of scores have been published with the indication that "Q-scores represent the ability to think in quantitative terms and the L-scores depend upon linguistic ability."[1]

Two forms for each general area were designed by designating the odd- and even-numbered items. The four forms were scored for a sample of 50 freshmen entering the Iowa State College in 1947.

The four interarea and the two intra-area correlations computed from scores of the four forms are shown in Table 17. The intra-area

Table 17

Inter- and Intra-Area Correlations of the ACE

Area		Quantitative		Linguistic	
		Odd	Even	Odd	Even
Q	Odd	-		.586	.679
	Even	.928		.565	.654
L	Odd			-	
	Even			.927	-

coefficients of reliability, stepped up by the Spearman-Brown formula were .963 for the Quantitative area and .962 for the Linguistic area.

---

[1]American Council on Education. Manual of instructions, psychological examination for college freshmen. Washington, D. C., American Council on Education. 1945. p. 7.

When the four interarea correlations and the two intra-area correlations were averaged by use of the function

$$1/2 \log_e \frac{1 + r}{1 - r}$$

$\bar{r}_A$ and $\bar{r}_W$ were found to be .623 and .928 respectively. Substituting $\bar{r}_A$ and $\bar{r}_W$ into the formula for testing the significance of departure from homogeneity

$$F = \frac{1 + \bar{r}_W - 2\bar{r}_A}{1 - r_W}$$

the formula became

$$F = \frac{1 + .928 - 2(.623)}{1 - .928} = \frac{.682}{.072} = 9.47$$

Consulting Snedecor's Table of F, with 49 and 49 degrees of freedom, the F-value of 9.47 is found to be highly significant. The hypothesis of homogeneity between the Q and L areas of the American Council on Education Psychological Examination is rejected since subjects are found to react differently to items within areas than to items between areas.

To obtain an indication of the extent to which the two areas of the American Council on Education Psychological Examination would be correlated if both areas were perfectly reliable, the average inter- and intra-area correlations were substituted into the proposed formula for correction for attenuation

$$r_{oo} = \sqrt{1 - \bar{r}_W^2 + \bar{r}_A^2}$$

and it became

$$r_{oo} = \sqrt{1 - (.928)^2 + (.623)^2} = .726$$

The coefficient of correlation corrected for attenuation in this manner suggests that there is some overlapping between the mental processes measured by this test, but that the true correlation is considerably less than unity.

B. Interest

The Wert-Myster Attitude Toward Farming Scale[1] yields scores for two areas, i.e., farming as a vocational choice and farming as a way of life. Although this scale was originally suggested for use as a measure of attitudes, norms have been established for this instrument with the prospect of using it as a measure of vocational interest.[2] The scale was selected for demonstration purposes as an interest test, since its scoring does not yield interdependent relationships as may be the case with interest tests composed of comparison-type items.

---

[1] A copy of this test is shown in the Appendix.

[2] Alonzo M. Myster. Construction and validation of a scale for the measurement of attitude toward farming. Unpublished Ph.D. Thesis. Ames, Iowa. Iowa State College Library. 1943.

The test consists of 46 items and responses are recorded on a five-point response scale. In computing the scores on this instrument the responses are weighted 0, 1, 2, 3, 4, the higher values indicating a preference toward farming. Nineteen items comprise the vocational area of the instrument and 27 items comprise the way of life area. The test is not timed, but the average length of time for responding is about 15 minutes for the 46 items.

The test was administered to 50 high school seniors and four scores were obtained by scoring the odd- and even-numbered items for the two areas. Spearman-Brown reliability coefficients were found to be .692 for the way of life area and .878 for the vocational area for this group of subjects.

The inter- and intra-area correlations for the four forms of the two areas are shown in Table 18.

Table 18

Inter- and Intra-Area Correlations of the Wert-Myster Scale

Area		Vocation		Way of Life	
		Odd	Even	Odd	Even
Vocation					
	Odd			.653	.517
	Even	.777		.700	.498
Life					
	Odd				
	Even			.521	

The average interarea correlation, $\bar{r}_A$, was found to be .599 and the average intra-area correlation, $\bar{r}_W$, was found to be .668.

When the foregoing values of $\bar{r}_A$ and $\bar{r}_W$ were substituted into the formula for testing the significance of departure from homogeneity, solution yielded an F-value of 1.42. With 49 and 49 degrees of freedom, this F-value is found to be nonsignificant. The responses of subjects to items within an area do not differ significantly from responses of subjects to items between areas, hence the null hypothesis is not rejected.

When the values of $\bar{r}_A$ and $\bar{r}_W$ were substituted into the proposed formula for correction for attenuation, the corrected correlation was found to be .955. The desirability of developing separate norms for each area of the Wert-Myster Attitude Toward Farming Scale would be questionable, since homogeneity was not disproved and since the coefficient of correlation corrected for attenuation approached unity.

C. Attitude

The Fritz Test of Cynicism which has been labeled "Practical Policy Test"[1] in an attempt to increase validity of subject responses, has been standardized on the basis of total score only.

---

[1]Charles O. Neidt. Analysis of college student reaction to the Fritz test of cynicism. Unpublished M. S. Thesis. Ames, Iowa. Iowa State College Library. 1947.

However, inspection of the 200 test items reveals that the items are subject to classification with respect to cynical attitude toward several institutions. To test the significance of departure from homogeneity among seven areas composed of test items concerning cynical attitudes toward different institutions, ten items[1] pertaining to each of seven more or less distinct areas were designated. These areas were called politics, business, labor, religion, marriage, education and wealth depending on the institution concerned.

The subject responds to each item along a four-point response scale and highly cynical responses are weighted 2, mildly cynical responses 1 and the two opposite responses are assigned 0 in computing the score. Data for this demonstration included scores for 400 Iowa State College students collected during 1945, 1946 and 1947.

Odd and even scores for each of the seven areas were obtained by weighting the responses to each of the ten items in an area. The remaining 130 items were not used in the analysis, although the method here described does not exclude their use. Ten items for each area were chosen primarily for convenience in computation.

Although the test is untimed, it is estimated that about six minutes per area, and about 45 minutes for the 70 items are required.

---

[1]Items comprising the seven areas are shown in the Appendix.

The Spearman-Brown reliability coefficients for the seven areas, for the 70 items and for the entire test of 200 items, are shown in Table 19. Inspection of the reliability coefficient

Table 19

Spearman-Brown Reliability Coefficients of the Fritz Test of Cynicism

Area	No. of Items	Reliability Coefficients
Politics	10	.748
Business	10	.793
Labor	10	.736
Religion	10	.336
Marriage	10	.719
Education	10	.628
Wealth	10	.764
Seven Areas	70	.877
Entire Test	200	.947

for the areas under consideration reveals that with the exception of the area of religion, the reliability coefficients are satisfactory for analysis, particularly when the small number of items in each area is considered.

Table 20

Inter- and Intra-Area Correlations of the Fritz Test of Cynicism

	Odd-Even	Business		Labor		Religion		Marriage		Education		Wealth	
		Odd	Even	Odd	Even	Odd	Even	Odd	Even	Odd	Even	Odd	Even
Politics	.598												
Odd		.498	.519	.415	.311	.231	.180	.384	.289	.219	.369	.583	.491
Even		.507	.471	.297	.241	.424	.254	.343	.236	.255	.373	.547	.515
Business	.657												
Odd				.324	.335	.243	.223	.260	.261	.241	.404	.631	.584
Even				.327	.354	.351	.447	.289	.288	.238	.345	.605	.534
Labor	.584												
Odd						.100	.149	.201	.165	.094	.174	.343	.269
Even						.058	.099	.156	.089	.008	.176	.262	.239
Religion	.224												
Odd								.180	.141	.203	.217	.244	.237
Even								.198	.077	.104	.133	.265	.260
Marriage	.554												
Odd										.194	.252	.452	.226
Even										.209	.275	.345	.240
Education	.458												
Odd												.372	.339
Even												.383	.342
Wealth	.618												
Odd													
Even													

The data needed for testing the significance of departure from homogeneity are shown in Table 20. Thus, for an over-all test of significance among the seven areas, $\bar{r}_W = .538$ (for all seven intra-area correlations), and $\bar{r}_A = .301$ (for all 84 interarea correlations).

Substituting, the formula then became

$$F = \frac{1 + .538 - 2(.301)}{1 - .538} = 2.04.$$

The numbers of degrees of freedom are 2394 and 399. By consulting Snedecor's Table of F, the departure from homogeneity is found to be highly significant.

Similarly, it is also possible to compute separately the F-values for each pair of areas. The results of such analysis are shown in Table 21. Inspection of this table reveals that each area is nonhomogeneous with each other area at the one per cent level of confidence with 399 and 399 degrees of freedom.

The proposed correction for attenuation also lends itself to the approaches used in the foregoing demonstration of the F-test for significance of departure from homogeneity, i.e., over-all analysis of the areas and comparison of each area with other areas.

Table 21

F-Values Between Areas of the Fritz Test of Cynicism

	Business	Labor	Religion	Marriage	Education	Wealth
Politics	1.69	2.34	1.54	2.24	1.97	1.37
Business		2.52	1.54	2.70	2.19	2.19
Labor			2.10	2.93	2.73	2.34
Religion				1.93	1.70	1.68
Marriage					2.12	2.30
Education						1.91

Required for significance 5% 1.19
1% 1.29

Substituting the values of $\bar{r}_A$ and $\bar{r}_W$ into the proposed formula for correction for attenuation for an over-all measure, the formula became

$$r_{oo} = \sqrt{1 - (.538)^2 + (301)^2} = .894$$

The results of obtaining the corrected correlation between each pair of areas are shown in Table 22. The greatest degree of overlapping between two areas was found for business and wealth and the least degree of relationship between labor and marriage.

Table 22

Proposed Corrections for Attenuation Between Areas of the Fritz Test of Cynicism

	Business	Labor	Religion	Marriage	Education	Wealth
Politics	.924	.867	.944	.854	.900	.957
Business		.852	.941	.840	.880	.970
Labor			.913	.836	.859	.869
Religion				.924	.945	.932
Marriage					.892	.870
Education						.904

D. Achievement

Appraisal of achievement is generally considered satisfactory only to the extent to which behavior indicated by course objectives has been evaluated. In the construction of achievement tests it is desirable to ascertain the degree to which course objectives must be broken down before homogeneous behavior is evidenced. Further, it is desirable to determine homogeneity among various objectives of a course. Since several approaches are usually available for evaluating pupil behavior, for example, essay- and objective-type tests, the degree of homogeneous behavior elicited by different types of tests is also appropriate for evaluation for homogeneity.

One example of each of the three foregoing circumstances was chosen for the following demonstration of the proposed formulas. A single objective which included different areas of subject matter was selected from conservation education. Evaluation among several objectives of a course in elementary educational statistics was chosen for the second demonstration, and the demonstration of essay- versus objective-type forms of a test was selected from home economics.

1. Single Objective.

A test of information of conservation of natural resources[1] was administered to 48 ninth-grade pupils. Although the test included only one objective, the recall of conservation information, the items from the test were classified into four areas: wildlife, forests, water and soils. Two scores, one for odd and one for even items, were obtained for each area. The test[2] consists of multiple-choice items with five choices and testing time is about one hour. Reliability coefficients for each area and for the entire test are shown in Table 23.

---

[1]Bernard F. Wievel. Attitude toward and knowledge of conservation possessed by students in Iowa high schools. Unpublished Ph. D. Thesis. Ames, Iowa. Iowa State College Library. 1947.

[2]A copy of the test is shown in the Appendix.

Table 23

Spearman-Brown Reliability Coefficients of the Conservation Information Test

Area	No. of Items	Reliability Coefficient
Wildlife	20	.304
Forests	20	.418
Water	15	.348
Soils	20	.665
Total Test	75	.728

The inter- and intra-area correlation coefficients are shown in Table 24. The intra-area correlations are too small for satisfactory reliability, but the range of scores here reported is considerably restricted.

Table 24

Inter- and Intra-Area Correlations of the Conservation Information Test

	Wildlife		Forests		Water		Soils	
	Odd	Even	Odd	Even	Odd	Even	Odd	Even
Wildlife								
Odd			.254	.287	.205	.255	.296	.336
Even	.179		.191	.230	.249	.087	.001	.435
Forests								
Odd					.452	.140	.107	.050
Even			.264		.241	.227	.346	.232
Water								
Odd							.132	.432
Even					.211		.113	.311
Soils								
Odd								
Even							.498	

When the correlations were averaged by the function

$$1/2 \log_e \frac{1 + r}{1 - r}$$

the values of $\bar{r}_A$ and $\bar{r}_W$ were found to equal .237 and .288 respectively. Substituting these values into the formula for testing the significance of departure from homogeneity, F was found to be 1.14. The numbers of degrees of freedom, (N-1) (k-1) and (N-1) are 141 and 47. An inspection of the Snedecor Table of F reveals nonsignificance. The hypothesis of homogeneity among areas is not

rejected. Pupils do not react significantly different to the four areas than they do to individual items within areas.

In Table 25 are shown the F-values which resulted when each

Table 25

F-Values Between Areas of the Conservation Information Test

	Forests	Water	Soils
Wildlife	.959	.99	1.23
Forests		.92	1.66*
Water			1.35

Required for significance 5% 1.62
1% 2.00

area was compared with every other area. With 47 and 47 degrees of freedom all except one F-value are found to be nonsignificant. The F-value between soils and water areas of the test is significant at the five per cent level.

Substituting the values of $\bar{r}_A$ and $\bar{r}_W$ into the proposed formula for correction for attenuation, the corrected correlation coefficient among all four areas was found to be .986.

The coefficients of correlation corrected for attenuation with the proposed formula for each pair of areas are shown in Table 26. With the exception of the correlation between forests and soils, which yielded the largest F-value, all correlations are larger than .950.

Table 26

Proposed Corrections for Attenuation Between Areas of the Conservation Information Test

	Forests	Water	Soil
Wildlife	1.00	1.00	.976
Forests		1.01	.941
Water			.965

The nonsignificant F-values and the high correlations for the areas in this test suggest that recall of information concerning natural resources represents a relatively homogeneous objective and that evaluation of this objective can be accomplished without constructing and scoring separate evaluation instruments for each area with which this objective is concerned.

2. Several Objectives.

Data concerning evaluation of achievement in several areas of elementary educational statistics were available for 22 Iowa State

College graduate students. Test scores were assembled for three of the major objectives of the course, i.e., information, interpretation and computation.

The informational objective included recall of facts, principles and formulas used in the elementary course. Interpretation dealt with ability to critically appraise the inference drawn in statistical investigations which were new to the student. Computation involved the solution of statistical problems with the use of the text and previous classroom notes.

Scores were obtained separately for the odd and even items for the tests covering each of the three objectives. Spearman-Brown coefficients of reliability and approximate time allotted for the testing of each objective are shown in Table 27.

Table 27

Spearman-Brown Reliability Coefficients of Tests Over Three Objectives in Statistics

Objectives	Testing Time	Reliability Coefficients
Information	30 minutes	.670
Interpretation	45 minutes	.765
Computation	3 hours	.868

Inter- and intra-area correlations necessary for substitution into the formula for testing the significance of departure from homogeneity are shown in Table 28. For an over-all test among the three objectives the F-value was found to be 2.55. With 42 and 21 degrees of freedom this F-value is found to be significant, but not highly so; although the value found approaches the value of 2.62 required for significance at the one per cent level. Students react

Table 28

Inter- and Intra-Area Correlations of Tests Over Three Objectives in Statistics

Objective	Information		Interpretation		Computation	
	Odd	Even	Odd	Even	Odd	Even
Information						
Odd			.611	.212	.555	.457
Even	.504		.288	.163	.179	.123
Interpretation						
Odd					.614	.370
Even			.620		.347	.395
Computation						
Odd						
Even					.775	

differently to items within areas or objectives than to items among areas or objectives. Hence, the hypothesis of homogeneity among objectives is rejected. Separate evaluation for each objective in elementary educational statistics is suggested.

The results of testing each pair of objectives are shown in Table 29. The F-value between interpretation and information is slightly lower than that required for five per cent significance and the other two approach significance at the one per cent level.

Table 29

F-Values Between Tests Over Three Objectives in Statistics

Objective	Interpretation	Computation
Information	2.06	2.86*
Interpretation		2.82

Required for significance 5% 2.07
1% 2.87

Computation of the correlation coefficient corrected for attenuation with the use of the proposed formula yielded a coefficient of .849 for the over-all analysis. The corrected correlation coefficients between each pair of areas are shown in Table 30. The relative overlapping for each pair of areas yields

Table 30

Proposed Corrections for Attenuation Between Tests Over Three Objectives in Statistics

Objective	Interpretation	Computation
Information	.890	.825
Interpretation		.833

the same ranking as was suggested by the relative size of the F-values in Table 29.

Further analysis of appraisal of these objectives would be appropriate because of the small number of cases here considered. However, emphasis is here being placed upon demonstration of the application of the formulas rather than analysis of the testing situation.

3. Validation of Type of Test.

Even though the course objectives to be evaluated with a test have been designated, there still remains the selection of the type of test to be constructed. In comparing two or more types of tests, which have been designed to measure the same behavior, for example, essay and objective types, it is appropriate to evaluate the degree to which the two types of the instrument measure homogeneous behavior. If the types of test can be scored by areas, these areas for each test can be evaluated independently. If nonhomogeneity among the areas within each type of test is indicated, the desirability of scoring each area separately is suggested.

Objective and essay types of the Kowitz Color Choice Test for Ninth Grade Home Economics Students[1] were administered to 50 ninth grade girls. Each type of the test was composed of two areas: area one measured the understanding of generalizations necessary for recognition of qualities of color, and area two contained test items in which problems were presented for solution. In the essay-type test the students were asked to write the qualities of colors and the generalizations involved and to write the solutions to the problems and provide their supporting reasons, whereas in the objective type of the test, the qualities of the colors and the solutions and reasons were checked from lists which were given to the students. For identification these areas will be designated as Essay I, Essay II and Objective I, Objective II. The essay type tests were scored jointly by two authorities in the field. An odd and an even score were obtained for each area for each subject. The following Spearman-Brown coefficients of reliability were obtained:

Essay I	.382
Essay II	.773
Objective I	.393
Objective II	.490

---

[1]Agnes E. Kowitz. Instruments for evaluation of home economics education: III. Test of application of generalizations to color choices of clothing for ninth grade girls. Unpublished M. S. Thesis. Ames, Iowa. Iowa State College Library. 1945.

The estimated time of administration for the two types of tests was 60 minutes for the objective and 90 minutes for the essay. The essay-type test was administered first.

The odd items of Essay I and the odd items of Objective I, etc., described the same test situation. For this reason in computing the inter- and intra-area correlations shown in Table 31, identical items were not correlated.

Table 31

Inter- and Intra-Area Correlations of Two Types of the Kowitz Achievement Test

	Essay				Objective		
	Area 1		Area 2		Area 1		Area 2
	Even	Odd	Even	Odd	Even	Odd	Even
Essay							
Odd	.236	.326	.203		.300	.088	.111
Even		-.002	-.224	.3615		-.112	.124
Odd			.627	.390	.053		.372
Even				.308	.072	.368	
Objective							
Odd					.244	.438	.280
Even						-.091	.240
Odd							.324
Even							

When the correlation coefficients were averaged by use of the function

$$1/2 \log_e \frac{1 + r}{1 - r}$$

$\bar{r}_A$ was found to be .185 and $\bar{r}_W$ was found to be .372 for the four intra-area correlations and for the 20 interarea correlations.

Substituting the values of $\bar{r}_A$ and $\bar{r}_W$ into the formula for testing the significance of departure from homogeneity and solving, an F-value of 1.60 for the over-all test of significance was found. With 49 and 147 degrees of freedom, this value does not reach significance at the five per cent level of confidence.

When the values of $\bar{r}_A$ and $\bar{r}_W$ were substitutded into the formula for correction for attenuation, the corrected coefficient was found to be .946 for all areas.

In Table 32 are shown the F-values for testing the significance of departure from homogeneity for each pair of areas. With 49 and 49

Table 32

F-values Between Areas of Two Types of the Kowitz Achievement Test

	Essay		Objective	
	Area 1	Area 2	Area 1	Area 2
Essay				
Area 1		2.28**		1.63*
Area 2			1.90*	
Objective				
Area 1				1.17
Area 2				

Required for significance 5% 1.61
1% 1.97

degrees of freedom, an F-value of 1.61 is required for significance

at the five per cent level and 1.97 at the one per cent level. Nonhomogeneous behavior is indicated between the two areas of the essay test, whereas homogeneous behavior is indicated for the corresponding areas of the objective test. Apparently naming colors and qualities of colors as was required in Essay I elicits a different type of behavior when compared to furnishing solutions to problems and supporting reasons in the choice of colors when the items are in essay type. When these same items are presented in matching form with supporting reasons to be checked, the behavior apparently becomes more homogeneous. The foregoing finding suggests that greater consideration must be given to the specific type of item when essay type tests are used than when objective type tests are used, particularly, if the different areas of the tests are to be combined in obtaining total scores for the tests.

In Table 33 are shown the correlations between each appropriate pair of areas after they have been corrected for attenuation with the use of the proposed formula. As in the F-test, Essay I and Essay II reflected the least homogeneous behavior when pairs of areas were compared.

Table 33

Proposed Corrections for Attenuation Between Areas
Two Types of the Kowitz Achievement Test

	Essay		Objective	
	Area 1	Area 2	Area 1	Area 2
Essay				
Area 1		.898		.961
Area 2			.915	
Objective				
Area 1				.987
Area 2				

V. SUMMARY

When total scores on psychological tests are obtained by combining numerical values assigned to individual test items, consideration must be given to the homogeneity of behavior represented by the items, if prediction of performance from such scores is to be accurate.

In educational evaluation the extent to which each objective must be subdivided before appraisal of instruction can be undertaken must also be considered. The testing of homogeneity among differing educational objectives is also appropriate as well as the determination of differences in behavior elicited by differing types of tests involving the same course material.

To fulfill the need for testing the homogeneity of separately-evaluated behavior characteristics, two formulas were developed. The first formula, F-test for significance of departure from homogeneity, was proposed to evaluate the null hypothesis: there is no significant difference in the manner in which individuals react to test items between or among areas of a test than to test items within areas of a test.

The formula is

$$F = \frac{1 + \bar{r}_W - 2\,\bar{r}_A}{1 - \bar{r}_W}$$

where $\bar{r}_W$ = average intra-area (odd-even) coefficient of correlation and $\bar{r}_A$ = average interarea coefficient of correlation and both averages are obtained with the use of the function

$$1/2 \log_e \frac{1 + r}{1 - r}.$$

The degrees of freedom for determining the significance of the F-value are (N-1)(k-1) and (N-1), where N = number of subjects and k = number of areas in the test.

The second formula, a correction for attenuation, was prepared as a positive indication of homogeneity. The formula is

$$r_{oo} = \sqrt{1 - \bar{r}_W^2 + \bar{r}_A^2}$$

where the notation is the same as for the foregoing formula.

To justify the use of the proposed formula for correction for attenuation and to compare the efficiency of correction of the proposed formula with the efficiency of four previously-developed formulas for correction for attenuation, three empirical demonstrations were designed.

In the first demonstration two perfectly-correlated hypothetical distributions of 20 values each were designated. These distributions were identified as "Basic$_1$" and "Basic$_2$". Four "test forms" were then developed, two from "Basic$_1$" and two from "Basic$_2$" by varying the size of the values in the basic distributions 0 or 1 at random.

Thus the true correlation was reduced from unity only because of the presence of "variable errors". This procedure was repeated ten times so as to result in ten applications of the formulas to four sets of test forms of 20 values and one application to a combined set of four test forms of 200 values. The five formulas for correction for attenuation under consideration were then applied. All formulas were found to be equally satisfactory in removing the effect of variable errors in the case of approximately equal means and variances in the test forms.

In the second demonstration the two test forms developed from "Basic$_1$" were obtained in the same manner as in the first demonstration. The two test forms developed from "Basic$_2$", however, were obtained by varying the basic distribution 0 to 9 at random. Thus one pair of test forms had a high correlation or reliability and the other a considerably lower correlation. As in the foregoing demonstration the procedure was repeated ten times. Because of a negative correlation in one of the ten sets of four test forms, all but the proposed formula were inappropriate for that set of four test forms. Although the true correlation in this demonstration was less than unity, all but the proposed formula yielded estimates of the true correlation higher than unity in one or more of the eleven applications. Estimates of the true correlation obtained with the use of the proposed formula were also more uniform than estimates obtained with the use of any one of the other four formulas.

In the third demonstration unlike basic distributions of 20 values each were used from which to develop the test forms. This procedure resulted in greater variability in the test forms than in either of the two preceding demonstrations. Again ten sets of four test forms based on 20 values and a combined set of four test forms based on 200 values were used. Estimates of the true correlation were obtained by applying the five formulas for correction for attenuation. A negative correlation in one of the sets of test forms rendered the four previously-existing formulas inappropriate for that set of test forms. As in the second demonstration, all except the proposed formula yielded estimates greater than unity in some cases. Greater variability among the estimates obtained with the use of the four previously-existing formulas was noted than among the estimates obtained with the proposed formula.

To demonstrate the use of the formula for testing the significance of departure from homogeneity and the proposed formula for correction for attenuation, several actual test situations were selected. The two formulas were applied to the results of an aptitude test, an interest test and an attitude test. Achievement testing was divided into a test situation involving a single course objective, one involving several course objectives each covered by one or more separate tests, and objective and essay types of a test involving the same course material.

Each test situation was chosen because of its appropriateness and emphasis was placed upon demonstrating the techniques rather than upon investigating the characteristics of the specific tests used.

It was found that the quantitative and linguistic areas of the American Council on Education Psychological Examination for College Freshmen elicited different types of behavior from a sample of 50 college freshmen.

The two areas of a scale to measure interest in farming, i.e., farming as a vocation and farming as a way of life, were found to elicit homogeneous behavior from 50 high school seniors.

Applying the technique to seven areas of the Frits Test of Cynicism, it was found that 400 college students reflected nonhomogeneous behavior in the cynical attitudes they expressed toward politics, labor, religion, marriage, education and wealth.

In a test involving recall of information concerning four areas of conservation, i.e., wildlife, forests, water and soils, homogeneous behavior was reflected among these four areas when 48 ninth-grade pupils were used as subjects.

When the formulas were applied to tests covering three objectives in a course in elementary statistics, recall of information, interpretation of statistical inference and computation, nonhomogeneous behavior was found. This group included only 22 graduate students as subjects, however.

Objective and essay-types of a test involving ability to apply generalizations concerning color and recall of color information were given to 50 ninth-grade home economics students. When the formulas were applied to the types of test concerned, it was found that the behavior elicited by the two areas of the objective test was homogeneous, whereas the behavior elicited by the two areas when the test was administered in essay form was nonhomogeneous.

In view of the results from applying the technique for testing the homogeneity of separately-evaluated behavior characteristics to hypothetical and actual test situations in this investigation, the technique is proposed for use in psychological testing and educational evaluation.

VI. SELECTED BIBLIOGRAPHY

American Council on Education. Manual of instruction, psychological examination for college freshmen. Washington, D. C., American Council on Education. 1945.

Barthelmess, H. M. The validity of intelligence test elements. Teachers College, Columbia University, Contributions to education. no. 505. 1931.

Brogden, Hubert E. Variation in test validity with variation in the distribution of item difficulties, number of items, and degree of their intercorrelation. Psychometrika. 11:197-214. 1946.

Brown, William. Some experimental results in correlation. Sixth International Congress of Psychology, Geneva, 1909. Proceedings p. 571-576. 1910.

Brown, William and Godfrey H. Thompson. The essentials of mental measurement. 4th ed. Cambridge. Cambridge University Press. 1940.

Carroll, John B. The effect of difficulty and chance success on correlations between items or between tests. Psychometrika. 10:1-19. 1945.

Clark, E. L. A method of evaluating the units of a test. Journal of Educational Psychology. 19:263-265. 1928.

Cook, W. W. The measurement of general spelling ability involving controlled comparisons between techniques. University of Iowa Studies in Education. Vol. 6: no. 6. pp. 38-59. 1932.

Cronbach, Lee J. Test "reliability": its meaning and determination. Psychometrika. 12:1-16. 1947.

Cureton, E. E. Quantitative psychology as a rational science. Psychometrika. 11:191-196. 1946.

Cureton, E. E. and J. W. Dunlap. Spearman's correction for attenuation and its probable error. American Journal of Psychology. 42:235-245. 1930.

Dunlap, J. W. Comparable tests and reliability. Journal of Educational Psychology. 24:442-453. 1933.

Festinger, Leon. The treatment of qualitative data by "scale analysis". Psychological Bulletin. 44:149-161.

Flanagan, J. C. General considerations in the selection of test items, etc. Journal of Educational Psychology. 30:674-680. 1939.

Goodenough, W. H. A technique for scale analysis. Educational and Psychological Measurement. 4:179-190. 1944.

Guttman, Louis. A basis for scaling qualitative data. American Sociological Review. 9:139-150. 1944.

Guttman, Louis. The Cornell technique for scale and intensity analysis. Educational and Psychological Measurement. 7:247-279. 1947.

Guttman, Louis and Edward A. Suchman. Intensity and a zero point for attitude analysis. American Sociological Review. 12:57-67. 1947.

Guilford, J. P. Psychometric methods. New York. McGraw-Hill Book Company. 1936.

Gulliksen, Harold. The relation of item difficulty and inter-item correlation to test variance and reliability. Psychometrika. 10:79-91. 1945.

Horst, P. Increasing the efficiency of selection tests. Personnel Journal. 12:254-259. 1934.

Horst, P. Item analysis by the method of successive residuals. Journal of Experimental Education. 2:245-263. 1934.

Hull, C. L. Aptitude testing. New York. World Book Company. 1928.

Jackson, R. W. B. Note on the relationship between internal consistency and test-retest estimates of the reliability of a test. Psychometrika. 7:157-164. 1942.

Kelly, T. L. Statistical method. New York. The MacMillan Company. 1923.

Kowitz, Agnes E. Instruments for evaluation of home economics education: III. Test of application of generalizations to color choices of clothing for ninth grade girls. Unpublished M. S. Thesis. Ames, Iowa. Iowa State College Library. 1945.

Lentz, T. F., B. Hirshstein, and F. H. Finch. Evaluation of methods of evaluating test items. Journal of Educational Psychology. 23:344-350. 1932.

Loevinger, Jane. A systematic approach to the construction and evaluation of tests of ability. Psychological Monographs. Vol. 61: no. 4. 1947.

Loevinger, Jane. The technic of homogeneous tests compared with some aspects of "scale analysis" and factor analysis. Psychological Bulletin. 45:507-529. 1948.

Long, J. A. Improved overlapping methods for determining validities of test items. Journal of Experimental Education. 2:264-267. 1934.

Lorge, I. The Thurstone attitude scales: I. Reliability and consistency of rejection and acceptance. Journal of Social Psychology. 10:187-198. 1939.

McNemar, Quinn. Opinion-attitude methodology. Psychological Bulletin. 43:289-374. 1946.

Myster, Alonzo M. Construction and validation of a scale for the measurement of attitude toward farming. Unpublished Ph.D. Thesis. Ames, Iowa. Iowa State College Library. 1943.

Neidt, Charles O. Analysis of college student reaction to the Fritz test of cynicism. Unpublished M. S. Thesis. Ames, Iowa. Iowa State College Library. 1947.

Paterson, D. G. Preparation and use of new-type examinations. New York. World Book Company. 1926.

Peters, Charles C. and Walter R. VanVoorhis. Statistical procedures and their mathematical bases. New York. McGraw-Hill Book Company. 1940.

Rider, Paul R. Modern statistical methods. New York. John Wiley and Sons. 1939.

Snedecor, George W. Statistical methods. 4th ed. Ames, Iowa. Iowa State College Press. 1946.

Spearman, C. The proof and measurement of association between two things. American Journal of Psychology. 15:72-101. 1904.

Spearman, C. Demonstration of formulae for true measurement of correlation. American Journal of Psychology. 18:161-169. 1907.

Spearman, C. Correlation calculated from faulty data. British Journal of Psychology. 3:271-295. 1910.

Thurstone, L. L and E. J. Chave. The measurement of attitude. Chicago. The University of Chicago Press. 1929.

Tucker, Ledyard R. Maximum validity of a test with equivalent items. Psychometrika. 11:1-13. 1946.

Uhrbrock, R. S. and M. W. Richardson. Item analysis. Personnel Journal. 12:141-154. 1933.

Vincent, L. A study of intelligence test elements. Teachers College, Columbia University. Contributions to Education no. 152. 1924.

Wievel, Bernard F. Attitude toward and knowledge of conservation possessed by students in Iowa high schools. Unpublished Ph.D. Thesis. Ames, Iowa. Iowa State College Library. 1947.

Yule, G. Udny. An introduction to the theory of statistics. 8th ed. London. Charles Griffen and Company. 1927.

VII. APPENDIX

ATTITUDE TOWARD FARMING

Alonzo M. Myster-James E. Wert

Name:_____Age:____Residence:(encircle one)Town Farm

School:_____Class:(encircle one) Fr So Jr Sr

Do you intend to be a farmer? (encircle one) Yes No

What are your opinions of the following statements? Your answer is correct if it expresses your true opinion. This is not a test and you are not to be graded. DO NOT OMIT ANY ITEM. In each case encircle the letter or letters which represent your own ideas about each statement.

SA-strongly agree; A-agree; U-undecided; D-disagree; SD-strongly disagree

1.	Farming is a pleasant vocation.	SA A U D SD
2.	I can't say I'm wild about farming.	SA A U D SD
3.	Farm work is drudgery.	SA A U D SD
4.	I would enjoy working with plants and animals.	SA A U D SD
5.	A farmer has more worries than do persons engaged in most other occupations.	SA A U D SD
6.	As a life's work, farming would be terrible.	SA A U D SD
7.	Work on the farm is really enjoyable.	SA A U D SD
8.	The disadvantages of farming outweigh its advantages.	SA A U D SD
9.	Farming requires less intelligence than most other occupations.	SA A U D SD
10.	The vocation of farming has its drawbacks, but I like it.	SA A U D SD
11.	Farming involves too many distasteful tasks.	SA A U D SD
12.	I have never wanted to be a farmer.	SA A U D SD
13.	Farming is fascinating work.	SA A U D SD
14.	Farming requires less education than most other vocations.	SA A U D SD
15.	I like farming well enough to make it my life's work.	SA A U D SD
16.	Farming is a monotonous vocation.	SA A U D SD
17.	I have always wanted to be a farmer.	SA A U D SD
18.	Living on a farm is just too much hard work.	SA A U D SD
19.	The advantages of farming far outweigh the disadvantages.	SA A U D SD
20.	Farming is uninteresting work.	SA A U D SD
21.	The average farmer enjoys farming more than I would.	SA A U D SD

(over)

22.	Everything considered, I could be happier farming than engaging in any other occupation.	SA	A	U	D	SD
23.	Farming tends to isolate a person from the rest of the world.	SA	A	U	D	SD
24.	Farming yields less satisfaction than most other occupations.	SA	A	U	D	SD
25.	The farm is the best place for young people.	SA	A	U	D	SD
26.	About all that can be said for farming is that it furnishes a means of existing.	SA	A	U	D	SD
27.	I could be satisfied to spend my life on the farm.	SA	A	U	D	SD
28.	Social well-being is impossible of attainment on the farm.	SA	A	U	D	SD
29.	Farming would give me a great deal of pleasure.	SA	A	U	D	SD
30.	I feel that I would be doing my children an injustice by living on a farm.	SA	A	U	D	SD
31.	I dislike the farm with its many inconveniences.	SA	A	U	D	SD
32.	Farming yields more satisfaction than most other vocations.	SA	A	U	D	SD
33.	Farming offers insufficient opportunity for contacts with people	SA	A	U	D	SD
34.	I feel that farm families as a whole are an interesting class of people.	SA	A	U	D	SD
35.	Farming has more social disadvantages than most other occupations.	SA	A	U	D	SD
36.	Farming deadens a person's ambitions.	SA	A	U	D	SD
37.	I like farming as a way of life.	SA	A	U	D	SD
38.	On an average, the standard of living of farmers is below that of other persons in the United States.	SA	A	U	D	SD
39.	The farm is wonderful place to live.	SA	A	U	D	SD
40.	Farming tends to reduce one's social standing.	SA	A	U	D	SD
41.	For me, life on a farm would be extremely dull.	SA	A	U	D	SD
42.	The farmer enjoys many satisfactions which the average person never realizes.	SA	A	U	D	SD
43.	I think I would have less fun living on a farm than in town.	SA	A	U	D	SD
44.	Farming deprives one's children of an adequate education.	SA	A	U	D	SD
45.	The independence of farm life appeals to me.	SA	A	U	D	SD

TEST ITEMS COMPRISING THE AREA OF POLITICS

(Fritz Test of Cynicism)

ODD TESTING ITEMS	CYNICAL RESPONSE
Politics is a racket. To call it a form of "legalized crookedness" would be a pretty good description.	Agreement
It is difficult for me to imagine an honest politician. He may mean very well when he starts out, but it is probable that he will not stay that way.	Agreement
You can get yourself elected to almost any public office if you have enough money.	Agreement
All you need is the "gift of gab" in order to be elected to an office in the United States.	Agreement
Although elections in America are democratic, the people do not have as much voice in the government as they think they do.	Agreement
EVEN TESTING ITEMS	
I would say that more than 90% of the people would not hesitate to be more or less "crooked" if they were absolutely certain that no one would find out about it.	Agreement
I think it is very near the truth to say that an honest man will not run for public office.	Agreement
I would say that perhaps as much as half of our tax money finds its way into the pockets of grafters.	Agreement
They say that all men are equal before the law, but I don't believe it.	Agreement
Every man in public office has a good possibility of becoming a "crook".	Agreement

TEST ITEMS COMPRISING THE AREA OF BUSINESS

ODD TESTING ITEMS	CYNICAL RESPONSE
The average individual who works for some big corporation is nothing more than a slave.	Agreement
Although the average big industrial corporation renders valuable service, you certainly pay for it plenty.	Agreement
99% of big business corporations will use bribery, lawyer's tricks and everything else to make money.	Agreement
The public be hanged. Business is business.	Agreement
You can depend on it that business will charge all the public will stand.	Agreement
EVEN TESTING ITEMS	
The average big industrial corporation is more humane, honest and public spirited than most people think.	Disagreement
For the most part, the little businessman does not have a chance because the big corporations will run him out of business.	Agreement
The average big industrial corporation certainly does "fleece" the public -- just a bunch of "crooks".	Agreement
For public protection, it is a mighty good thing that the government controls big business as much as it does.	Agreement
99% of big business corporations are out to make all the money they can.	Agreement

TEST ITEMS COMPRISING THE AREA OF RELIGION

ODD TESTING ITEMS	CYNICAL RESPONSE
Going to church will cause an individual to treat his fellow men with kindness, justice and mercy.	Disagreement
Some of the biggest hypocrites are found in church.	Agreement
Religion is the "backbone" of any nation.	Disagreement
I think the stories such as "Jonah and the Whale", "Noah's Ark" and many others are just stories. They did not actually take place.	Agreement
People who won't play cards on Sunday are narrow-minded.	Agreement
EVEN TESTING ITEMS	
Clergymen, with very few exceptions, get through life pretty easily, that is to say, they have what we would ordinarily call a "soft" job.	Agreement
I would say that nearly all missionaries are sincere about their work.	Disagreement
Anyone who takes religion very seriously is lacking in sophistication, that is, he may be considered a bit gullible.	Agreement
How I want to spend my time on Sunday is nobody's business.	Agreement
It is just pure foolishness to deny yourself something during Lent.	Agreement

TEST ITEMS COMPRISING THE AREA OF LABOR

ODD TESTING ITEMS	CYNICAL RESPONSE
Labor unions are out to get all they can for themselves.	Agreement
I just don't see how a man can belong to a labor union and still be a good citizen.	Agreement
Labor leaders, with very few exceptions, are sincere, and really have the welfare of the workers in mind.	Disagreement
I think that most laborers are just radicals.	Agreement
Laborers are paid what they are worth. If they aren't paid more, it is because they aren't worth any more.	Agreement
EVEN TESTING ITEMS	
Men who join labor unions are dangerous.	Agreement
Most laborers are "hidebound" and conservative.	Agreement
Laborers for the most part, are just too dumb to know any better. If they don't get along, it is due to their own stupidity.	Agreement
Labor leaders are just a bunch of racketeers. They are getting all they can out of it for themselves.	Agreement
Laborers should be paid a higher wage. They deserve more than they are getting.	Disagreement

TEST ITEMS COMPRISING THE AREA OF MARRIAGE

ODD TESTING ITEMS	CYNICAL RESPONSE
The chances of meeting and falling in love with just the right person so that you will have a truly happy marriage are really very small.	Agreement
"Falling in love" is just a big trick that nature plays on people.	Agreement
Most women marry because they think it will be easier than getting a job and working.	Agreement
Marriage is just another way of saying "Battlefield".	Agreement
I believe that at least 90% of the girls would rather marry a poor boy whom they love than a rich man whom they do not love.	Disagreement
EVEN TESTING ITEMS	
There is really considerable justification for the average man to refer to his wife as a "ball and chain".	Agreement
Marriage is just another way of "jumping out of the frying pan into the fire". Not to marry may be unsatisfactory, but certainly marriage is worse.	Agreement
About the only good reason for getting married is that society thinks you are abnormal if you don't.	Agreement
The average man looks upon a wife as a servant.	Agreement
Most mothers try to make "sissies" out of their sons.	Agreement

TEST ITEMS COMPRISING THE AREA OF EDUCATION

ODD TESTING ITEMS	CYNICAL RESPONSE
Most college professors are impractical, with very little appreciation of the real, everyday world.	Agreement
I think it is certainly true that you don't learn how to think in college.	Agreement
Most people who graduate from college are not really educated.	Agreement
Colleges and universities put too much money into magnificent buildings and not enough in straight "honest-to-goodness" education.	Agreement
The chief purpose of football teams and basketball teams is to build better character.	Disagreement
EVEN TESTING ITEMS	
About 90% of our education at the present time is a sheer waste of time.	Agreement
Getting a higher education is next thing to a dead loss of time.	Agreement
Football teams are mostly for the purpose of advertising a college or university.	Agreement
The country is full of "educated fools" who have graduated from college.	Agreement
If it were not for the social life, college would be a complete waste of time.	Agreement

TEST ITEMS COMPRISING THE AREA OF WEALTH

ODD TESTING ITEMS	CYNICAL RESPONSE
If you have money, you can do just about anything you like.	Agreement
The average American is just a money-grabber.	Agreement
If you lend money to your friends and relatives, they will consider you just "an old Shylock" when you ask for its return.	Agreement
Most millionaires are just clever crooks who have taken advantage of the "breaks".	Agreement
I think that most millionaires have had to be "crooked" at least part of the time.	Agreement

EVEN TESTING ITEMS

Whenever I pass the estate of a very wealthy person (where I see a big house, servants, fine cars and all that) I just can't help thinking he must have been somewhat "crooked" to get it all.	Agreement
Great fortunes, with very few exceptions, have been built up by exploiting the poor.	Agreement
"Soak the rich" sounds like a good policy to me.	Agreement
I believe that most wealthy people have been at least as honest as the average person.	Disagreement
Money may be the "root of all evil", but give me more root and I'll take a chance on the evil.	Agreement

PERSONAL DATA QUESTIONNAIRE

1. Name..... 2. Age..... 3. Sex.....

4 Name of School..... 5. Grade.....

6. Most or all of my elementary work was done in a consolidated school.....; town school.....; country school.....

7. I live on a farm.....; in town.....

8. My favorite hobbies are....., ..., ...

9. My grades are above average.....; average.....; below average.....

10. Underline the courses THAT YOU HAVE TAKEN PRIOR TO THIS YEAR in the following list. Underline the activities in which YOU HAVE PARTICIPATED OR ARE NOW PARTICIPATING.

Courses:		*Activities:*	
General Science	Biology	4-H Club	Bird Clubs
Nature Study	Zoology	FFA	Nature Camps
Natural Science	Botany	Summer Camp	Izaak Walton League
General Agriculture	Physics	Boy Scouts	
Vocational Agriculture	Civics	Girl Scouts	
Industrial Arts	Chemistry	Camp Fire Girls	
Home Economics	Geography	Conservation Clubs	
Physiology	Sociology	Nature Photography	

11. Indicate the extent to which you read the following magazines by placing a check mark in the proper column.

	Regularly	Occasionally	Never		Regularly	Occasionally	Never
Nature Magazine.....	.....	.....	.....	Natural History.....	.....	.....	.....
American Forests.....	.....	.....	.....	Outdoor America.....	.....	.....	.....
Audubon Magazine.....	.....	.....	.....	Outdoor Life.....	.....	.....	.....
Iowa Conservationist ...	.....	.....	.....	Sports Afield.....	.....	.....	.....

12. I listen to radio programs on conservation topics regularly.....; occasionally.....; never.....

13. I read books and bulletins on conservation regularly.....; occasionally.....; never.....

14. I read newspaper articles on conservation regularly.....; occasionally.....; never.....

15. I see motion pictures on conservation regularly.....; occasionally.....; never.....

ATTITUDE TOWARD CONSERVATION

What are your opinions of the following statements? Your answer is correct if it expresses your true opinion. This is not a test and you are not to be graded. DO NOT OMIT ANY ITEM. In each case encircle the letter or letters which represent your own ideas about each statement.

SA—Strongly agree; A—agree; U—undecided; D—disagree; SD—strongly disagree.

1. Even though the natural resources of the tropical countries are largely undeveloped there is an urgent need for conservation of temperate zone resources	SA	A	U	D	SD
2. The growth of our country will be retarded if we conserve our resources	SA	A	U	D	SD
3. A man should be able to use his land as he sees fit	SA	A	U	D	SD
4. The mining of all minerals should be controlled by the government	SA	A	U	D	SD
5. For a fair price the state should have the right to buy any lands for conservation purposes	SA	A	U	D	SD
6. Floods of the lower Mississippi should be controlled with levees	SA	A	U	D	SD
7. The killing of waterfowl should be limited to hunters who have means of retrieving them	SA	A	U	D	SD
8. Only funds obtained by sale of hunting and fishing licenses should be used for conservation of wildlife	SA	A	U	D	SD
9. The number of trees cut on privately owned woodlots should be controlled by the government	SA	A	U	D	SD
10. Soil erosion control should be left entirely to the landowner	SA	A	U	D	SD
11. Financial aid in conservation work makes the farmer too dependent on the government	SA	A	U	D	SD
12. It is wise for the government to spend large sums of money on erosion control	SA	A	U	D	SD
13. Soil erosion is a minor problem	SA	A	U	D	SD
14. Whether or not the government aids in the control of soil erosion is little concern of mine	SA	A	U	D	SD
15. We are needlessly and rapidly wasting our soil	SA	A	U	D	SD
16. In the interests of wildlife the polluted condition of our rivers and lakes should receive immediate attention	SA	A	U	D	SD
17. Recent scientific developments have made it unnecessary to conserve our natural resources	SA	A	U	D	SD
18. A continuous open season on the fox is entirely justified	SA	A	U	D	SD
19. Hunters and fishermen should have the last word in wildlife conservation	SA	A	U	D	SD
20. People interested in conservation tend to stand in the way of industrial progress	SA	A	U	D	SD
21. An alert conservation group is a necessity in every community	SA	A	U	D	SD
22. The schools should spend more time teaching conservation	SA	A	U	D	SD
23. There is an urgent need for nationwide public concern over conservation education	SA	A	U	D	SD
24. Conservation of natural resources is a dull subject	SA	A	U	D	SD
25. The average man has little need for education in conservation	SA	A	U	D	SD

KNOWLEDGE OF CONSERVATION

Directions: Each of the following questions is followed by three or four answers of which ONLY ONE IS CORRECT. Choose the one that is correct and put its number in the parentheses at the right.

Sample: Fireflies are—1. wasps 2. beetles 3. true flies 4. bees (2)
Fireflies are a kind of beetle so the number "2" is placed in the parentheses

1. The most numerous and valuable fur-bearer in Iowa is the—1. mink 2. red fox 3. muskrat 4. striped skunk ( )

2. The most hunted game mammal in the state is the—
1. fox squirrel 2. red fox 3. grey squirrel 4. cottontail rabbit ( )

3. The most numerous game bird in central and northern Iowa counties is the—
1. ruffed grouse 2. Hungarian partridge 3. bobwhite 4. ringnecked pheasant. ( )

4. The most numerous game bird in southern Iowa is the—
1. ruffed grouse 2. prairie chicken 3. bobwhite 4. ring-necked pheasant ( )

5. The food of the red fox consists principally of—
1. mice and rabbits 2. farmer's poultry 3. pheasants and quail 4. songbirds ( )

6. Pheasants are often blamed for damaging sprouting corn but the real culprits are more often—
1. ground squirrels 2. mink 3. bobwhites 4. grey squirrels ( )

7. The success of the Iowa State Conservation Commission's program depends primarily on the number of—
1. licenses sold 2. conservation officers in the field 3. scientists on the staff 4. people approving conservation policies ..... ( )

8. To improve hunting and fishing in Iowa the money received from hunting and fishing licenses is better spent for—
1. enlarging fish and game hatcheries 2. catching game law violators 3. purchasing game and fish from near-by states 4. habitat improvement..... ( )

9. The ability of either land or water to produce game or fish depends primarily on—
1. soil and water fertility 2. the weather 3. control of predators 4. cosmic forces ..... ( )

10. To improve bird shooting the State Conservation Commission should first—
1. increase quail production in northern Iowa 2. increase pheasant production in southern Iowa 3. aid each dominant species within its own range 4. restock with prairie chickens ..... ( )

11. Artificial restocking is a sound management practice when—
1. the public clamors for better hunting 2. hunters have difficulty in bagging wild game 3. conservation agencies have funds 4. game is not present in suitable areas ( )

12. In stocking farm ponds the two fishes which provide a good balance are the—
1. channel catfish and yellow bullhead 2. largemouth bass and the bluegill 3. smallmouth and largemouth bass 4. yellow bullhead and bluegill..... ( )

13. The factor which at present needs more attention in order to increase Iowa's pheasant population is—
1. food 2. predator control 3. roosting cover 4. nesting cover ..... ( )

14. The difficulty in providing good duck hunting for Iowa hunters is primarily due to the—
1. scarcity of ducks 2. warm fall weather 3. early freeze-up 4. lack of desirable places to hunt..... ( )

15. The real danger in having too long a pheasant hunting season is that—
1. too many cock birds will be killed 2. hunting accidents will increase 3. too many hen birds will be killed 4. hunters will expect a long season every year ..... ( )

16. In prairie country the plant type most suitable for wildlife and for erosion control is—
1. grass 2. weeds 3. trees 4. shrubs ..... ( )

17. The major aim of wildlife managers is to—
1. please the public 2. maintain wildlife populations at desirable levels 3. control predatory animals 4. enforce game laws ..... ( )

18. The number of major migratory bird routes or flyways in the United States is—
1. two 2. four 3. six 4. eight .....( )

19. The best trout fishing in Iowa is found in the—
1. northern lakes section 2. artificial lakes 3. southern streams 4. streams of the northeastern section ..... ( )

20. In general the most successful practice to improve fishing is to—
1. provide more fish hatcheries 2. restock the waters with large fish 3. limit the take of fish 4. improve water conditions to encourage natural reproduction ... ( )

21. In Iowa the banding or marking of birds and animals for scientific study is—
1. unlawful at all times 2. lawful provided a permit has been obtained 3. encouraged to promote student interest 4. permissable only during the open season.... ( )

22. The value of Iowa's fur crop could be greatly increased if—
1. trappers would use greater care in handling furs 2. trapping seasons were extended 3. more humane traps were used 4. the season on otters was opened.. ( )

23. Insects are—
1. of little economic value 2. in danger of becoming extinct because of DDT 3. seldom eaten by game birds 4. often eaten by skunks ..... ( )

24. Most Iowa snakes are—
1. beneficial and should be protected 2. of no importance 3. destructive of songbirds 4. dangerous and should be killed whenever possible ..... ( )

25. When hunting migratory waterfowl it is legal to use—
1. five live decoys 2. corn to attract the birds 3. artificial decoys 4. two live decoys plus five artificial ones ( )

26. The ratio of man-caused forest fires to all other causes is about—
1. 2:1 2. 3:1 3. 5:1 4. 9:1 ..... ( )

27. Rapid detection of forest fires is accomplished in the National Parks and Forests chiefly by—
1. a lookout system 2. tourist cooperation 3. airplane patrol 4. horse patrol ..... ( )

28. The outstanding need in protecting our forests from fire is—
1. an intensive educational program 2. more severe laws 3. fire lane construction 4. improved fire fighting practices ..... ( )

29. The trend in size of the average forest fire and the area burned over is—
1. generally upward 2. remaining fairly constant 3. generally downward ..... ( )

30. Heavy grazing of farm woodlands relative to forest reproduction is considered as—
1. helpful 2. very helpful 3. slightly harmful 4. very harmful ..... ( )

31. The U. S. Forest Service is a bureau of the—
1. Department of the Interior 2. Fish and Wildlife Service 3. Department of Agriculture 4. National Park Service ..... ( )

32. National Parks differ from the National Forests because in the former—
1. timber cutting is permitted 2. grazing is permitted 3. recreational use is prohibited 4. commercial use is forbidden ..... ( )

33. The largest single owner of forest lands in the country is the—
1. United States government 2. State of California 3. Weyerhauser Lumber Company 4. State of Washington ( )

34. Most of the forested area of Iowa is owned by the—
1. State 2. farmers 3. United States 4. lumber companies ..... ( )

35. The actual practice of scientific forestry in Iowa has been—
1. sadly neglected 2. accepted by woodland owners 3. proved unnecessary 4. proved of slight value..... ( )

36. In Iowa the percentage of the present timbered areas compared to that of the State's total area is about—
1. 7% 2. 17% 3. 27% 4. 37% ..... ( )

37. Woodlands on Iowa farms should be—
1. cleared and planted to crops 2. cut selectively as trees become mature 3. protected from all cutting 4. grazed enough to keep down tree reproduction..... ( )

38. The most abundant species of trees in Iowa are the—
1. oaks 2. maples 3. cottonwoods 4. walnuts ..... ( )

39. Wasteful use of our forests is more apt to occur from the practices of the—
1. Federal government 2. large lumber companies 3. small woodlot owners 4. small contractors..... ( )

40. The most important commercial forest area in the United States is the—
1. Northeastern Region 2. Plains Region 3. California Region 4. Southern Region ..... ( )

41. The mineral resources in the United States are—
1. evenly distributed 2. limited and exhaustible 3. unlimited 4. renewable..... ( )

42. In the United States there is—
1. no serious lack of any mineral 2. a lack of some unimportant minerals 3. a serious lack of coal, iron and oil 4. a serious lack of twenty or more industrial minerals ..... ( )

(TURN THE PAGE)

43. The most abundant mineral found in Iowa among the following is—
1. lead 2. coal 3. zinc 4. bauxite ..... ( )

44. Our known oil resources are—
1. sufficient to last 1000 years 2. more than one-half gone 3. more than three-fourths gone 4.practically inexhaustible ..... ( )

45. The most difficult part of a metal conservation program to have adopted in the United States is—
1. prevention of production wastes 2. the extensive use of substitutes 3. prevention of unessential use of metals 4. development of a national habit of saving..... ( )

46. The problem of water conservation—
1. is closely related to all fields of conservation 2. is unimportant in most Middle-western areas 3. is of less importance than wildlife conservation 4. has been solved in Iowa ..... ( )

47. Drainage of all Iowa's swamps and marshes—
1. should be undertaken to kill mosquitoes 2. should be undertaken to provide more cropland 3. should be undertaken to reduce muskrat populations 4. is not justified ..... ( )

48. The main problem in the conservation of water is—
1. to hurry its return to the sea 2. to distribute and use it often where needed 3. the construction of reservoirs 4. the drainage of swamplands ..... ( )

49. The loss of moisture by evaporation is greatest from—
1. barren soils 2. forest soils 3. grassland soils 4. cornfields ..... ( )

50. Loss of moisture from the ground through transpiration is greatest from—
1. barren soil 2. hardwood forests 3. pine forests 4. grasslands ..... ( )

51. The best water control method for making certain arid lands productive is to—
1. drain the areas 2. construct terraces 3. use dry land fallow 4. irrigate ..... ( )

52. Records show that floods in the United States are—
1. becoming more severe, more frequent and destructive 2. becoming less frequent, less severe and less destructive 3. not severe enough to cause much concern 4. becoming less frequent but more destructive ..... ( )

53. Stream pollution in Iowa—
1. must be tolerated if we are to have manufacturing 2. is practically nonexistent 3. exists but is not serious 4. costs the State many thousands of dollars ..... ( )

54. Heavy pollution in a stream destroys fish life by—
1. reducing the oxygen content 2. reducing the carbon dioxide content 3. preventing their free movement 4. encouraging growth of water plants..... ( )

55. The chief purpose of a shelterbelt of trees is to—
1. reduce wind velocity and conserve moisture 2. decrease moisture loss by transpiration 3. provide shade for growing crops ..... ( )

56. Another name often given to the top layer of soil is the—
1. A horizon 2. B horizon 3. C horizon 4. parent material ..... ( )

57. The productivity of a soil is usually measured in terms of its capacity for—
1. growing crops 2. retaining moisture 3. resisting erosion 4. releasing moisture ..... ( )

58. The original plant growth that covered Iowa was—
1. prairie grasses only 2. prairie grass with evergreen forests along the streams 3. hardwoods only 4. prairie grass with hardwoods along the streams ..... ( )

59. Photosynthesis is the photo-chemical reaction by which plants—
1. manufacture food materials 2. use food materials 3. give off carbon dioxide 4. give off water ..... ( )

60. The dark color of most Iowa streams is due primarily to—
1. natural color of the water 2. industrial pollution 3. slow, sluggish flow 4. excessive amounts of soil in the water ..... ( )

61. If water-loss or run-off occurs, it would be greatest from—
1. a burned-out woodland 2. a grazed pasture 3. a virgin woodland 4. a wheat field ..... ( )

62. Iowa's most valuable natural resource is her—
1. soil 2. water 3. wildlife 4. minerals ..... ( )

63. The number of chemical elements needed by plants to maintain vigorous growth is—
1. four 2. eight 3. ten 4. fourteen ..... ( )

64. Moderately uniform removal of surface soil is known as—
1. sheet erosion 2. gully erosion 3. waterfall erosion 4. wind erosion ..... ( )

65. During the 100 years that Iowa land has been under cultivation the part of the original topsoil that has been removed is—
1. none 2. one-fourth 3. one-third 4. three-fourths... ( )

66. The first step in launching a soil conservation program would be to—
1. make a study of soil conditions 2. plant permanent vegetation 3. devise a good crop rotation 4. build barriers against erosion ..... ( )

67. Accelerated or man-induced erosion is causing our top soil to be removed—
1. faster than it is being formed 2. at about the same rate it is being formed 3. at a slower rate than it is being formed ..... ( )

68. Iowa ranks first in the United States in value of farm products chiefly because of its—
1. long growing season 2. good farmers 3. fertile soil 4. agricultural education programs ..... ( )

69. The three elements most likely to be needed as fertilizers by Iowa soils are—
1. nitrogen, calcium and iron 2. calcium, phosphorus and iron 3. nitrogen, phosphorus and potassium 4. potassium, phosphorus and calcium..... ( )

70. A strip of close-growing, soil-holding grass planted on the contour is called a—
1. turning lane 2. grassed waterway 3. diversion ditch 4. buffer strip ..... ( )

71. Surface soils always have—
1. better structure than sub-soils 2. less pore space than sub-soils 3. less plant nutrients than sub-soils 4. less organic matter than sub-soils..... ( )

72. The chief reason for growing inoculated legumes such as sweet clover and alfalfa is to restore—
1. air to the soil 2. calcium to the soil 3 nitrogen to the soil 4. carbon dioxide to the soil ..... ( )

73. The best method used by farmers enabling them to farm wet soils is—
1. planting crops in rotation 2. installing a drainage system 3. plowing under crop residues 4. returning manure promptly ..... ( )

74. Freezing and thawing, wetting and drying of a soil—
1. has a detrimental effect on its structure 2. improves its structure 3. has no effect on its structure 4. increases its organic matter ..... ( )

75. The most practical way of improving soil structure is to—
1. add humus by growing grass 2. add humus by growing intertilled crops 3. control weed growth 4. contour the slopes ..... ( )

Prepared by Bernard Wievel
Iowa State College